THE UNIVERSITY
of EDINBURGH

Quantitative Genetic and Genomic Analyses of the Effect of Porcine Reproductive and Respiratory Syndrome (PRRS) Outbreaks on the Reproductive Performance of sows

Christopher Mark Orrett

THE UNIVERSITY *of* EDINBURGH

Royal (Dick) School of Veterinary Studies. Animal Genomics and Disease Resistance. PhD

The University of Edinburgh

2017

This Thesis is Dedicated to

Professor Stephen C. Bishop

and

Grandma Mavis Whittingham

"All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy. All work and no play makes Jack a dull boy."

(The Shining, 1980)

Declaration

I declare that this thesis is of my own composition; that the body of work contained herein is my own, except where stated otherwise by reference or acknowledgment; it has not been submitted in whole or in part for any other degree or qualification.

Christopher Mark Orrett

31st August 2017

Acknowledgements

Without the support of the late Prof. Stephen C. Bishop (1960-2015), this thesis wouldn't have been started, let alone finished. Science can sometimes lack in personality, yet Steve's unbridled cheeriness and dry wit would lighten even the most uninspiring of topics. Steve's love and in-depth knowledge of his subject combined with his enthusiasm and infectious personality was an inspiration for not just me, but for the many students and staff who had the pleasure of working with him. Many a Friday afternoon Steve and I would far exceed the one-hour supervisor meeting we had planned, attracting attention with the loud guffaws and heated debates emanating from his office. Thank you.

Special thanks must go to Prof Alan L Archibald and Dr Andrea Doeschl-Wilson both of whom have helped so much in the formulation of this thesis, under difficult circumstances. Andrea kindly agreed to assist in the supervision of this project following the tragic death of Stephen Bishop and her bright, fresh insight has made a valuable contribution. Alan has been a huge help, a stalwart of the Roslin Institute, who has the same dry, insightful-yet-modest character of Stephen which has been a source of great support. The input from the Genetics and Genomics team at Roslin has been excellent, but specific mention must be made for Dr Oswald Matika, Dr Valentina Riggio and Dr Riccardo Pong-Wong. Their expertise and assistance have been vial in enabling me to conduct this research. I am also indebted to the wider Roslin Institute but especially Professor Bernadette Dutia, Liz Archibald and Liz Brown.

Additional support was provided by PIC/Genus PLC who collaborated on aspects of this project, provided the data for analysis, funding and the opportunity to work alongside them in Hendersonville TN. Specifically Dr Joseph (Nader) Deeb, Dr David G McLaren, Dr Craig Lewis and Saad Sebbana.

This PhD received additional financial support from the Biotechnology and Biological Sciences Research Council (BBSRC) without whom, this research and indeed much of the UK's biological science research output, would not be possible.

The final thanks are always reserved for those dear people who mean so much. My Grandma Mavis Whittingham (1935-2017); without her love and support I would have wasted away many years ago and whose pride in this accomplishment means so much to me. A further special note must be made for the three people who enabled me to finish, maintain and start this PhD, whom I love, Peter James Cloutman, Amy-Curtis Hamilton and David Michael Barclay, and who continue to mean so much to me. There are so many people to consider including my peers at Roslin for the jolly banter, my wider circle of friends, Rev Jude Mitson, Neil Crandon and other family members such as my Grandad, Albert Whittingham and Auntie Karen. However, specifically, it is Stephen Bishop, Grandma, Peter, Amy, David with the addition my ever-loving parents to whom I owe the most.

Contents

Table of Figures

Table of Tables

List of Abstracts

Orrett, C.M., Matika, O., Archibald, A., Lewis, C.R.G., Mclaren, D., Deeb, N. & Bishop, S. (2013). Genetic of host response to infection with porcine reproductive and respiratory syndrome virus (PRRSv). Advances in Animal Bioscience (Proceedings of the British Society of Animal Science 2013, Nottingham). 4. p. 81.

Orrett, C.M., Deeb, N., Pong-Wong, R., Matika, O., Lewis, C.R.G., McLaren, D.G., Archibald, A. & Bishop, S. (2014). Regional Heritability Mapping of Production Traits in Epidemic Porcine Reproductive and Respiratory Syndrome. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. 2014, Vancouver, p. 100.

Abstract

Porcine Reproductive and Respiratory Syndrome (PRRS) is, globally, one of the costliest of diseases to the pig industry. Despite enormous efforts, methods such as vaccination strategies and herd management have failed to fully control the disease. Exploiting the genetic variation in host response could be included as part of a multifaceted approach to mitigate the devastating impact of this disease. Establishing the presence of genetic variation and its underlying genetic architecture are key to implementing genomic selection, which is considered a viable and safe long-term disease control strategy. This thesis explores the effect of natural PRRSV outbreaks on the reproductive performance of sows, and the underlying genetic influences on it.

Litter records were available from two farms, where Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) outbreaks had been confirmed using ELISA. One farm had full pedigree information, but for both farms 60K SNP genotypes were available. In both farms, performance records could be partitioned into an epidemic and non-epidemic phase using a previously established threshold method. The partitioning also identified a period of high reproductive failure not coinciding with a diagnosed PRRSV outbreak on one farm. This period was isolated and analysed separately.

Linear mixed models were used to explore both genetic and non-genetic factors contributing to differences in reproductive performance associated with the two phases. This analysis identified five disease indicator traits identified showing significant differences (>95% CI) in least squares means between the epidemic and non-epidemic phase. These were the number of mummified, stillborn, dead and alive piglets per litter and the fraction of the total born dead. Alternative statistical models that accounted for differences in the severity of the individual PRRSV outbreaks were also considered throughout. Despite differences in the

estimates associated with different models and farms, in general very low heritability estimates were obtained for these disease indicator traits during the non-epidemic phase, whereas the traits were found moderately heritable during the epidemic phase.

Two genome wide association analyses methods were used to explore the distribution of the genetic effects throughout the genome: Family-based Score Test for Association (FASTA) and Genome-wide Rapid Analysis using Mixed Model and Regression (GRAMMAR). In addition, regional associations were studied using Regional Heritability Mapping (RHM). Associations were then further characterised using Measured Genotype (MG) analyses.

Genome-wide significant associations were identified for five SNPs and one region. The regional association spans the region previously identified in an experimental challenge experiment of growing pigs, in association with viral load and weight gain. Different patterns of linkage disequilibrium (LD) are observed which may explain why this study and others failed to find single SNP effects at this location. One genome wide significant SNP on SSC15 was found between two previously identified SNPs associated with PRRSV mortality. Five further putative SNP associations are indicated by RHM and subsequent measured genotype analysis, two of which flank previously reported associations and indicate an epistatic effect, observed in several traits.

In summary, this study showed that reproductive performance of sow is considerably reduced during PRRSV outbreaks and the genetics of the sow significantly affects variance in survival and mortality. Several novel genomic regions associated with the reproductive performance of sows in the absence and during PRRSV outbreaks have been identified in this study. In addition to these, the results suggest the region on SSC4 previously associated with PRRSV viral load and weight gain may also affect foetal mortality. These results demonstrate the potential for genomic selection to be used to mitigate PRRSV related reproductive losses, the greatest financial exposure faced by the pig industry. In addition, RHM is directly shown to capture genetic variance, where single SNP methods fail to

identify an effect, highlighting the usefulness of this tool as a method to identify genomic regions with significant effect on production traits.

Lay Summary

Porcine Reproductive and Respiratory Syndrome (PRRS) is a highly infectious viral disease of pigs with devastating effects on pig production across the world. It can cause the death of piglets before and shortly after they are born. Traditional methods have failed to control the spread of the disease leading to considerable interest in breeding pigs that can better cope with the disease. This study explores the impact of PRRS outbreak on piglet survival, to what extent piglet survival rates of pregnant sows exposed to PRRS is heritable, and which region of the genome control these survival rates.

Information was available from two farms experiencing repeated outbreaks of PRRS. Measurements included records for piglet survival at birth and weaning together high density genomic information as well as pedigree information, from one of the farms only. Combining the data from the two farms gave more accurate estimates of genetic effects of sows exposed to PRRS on piglet survival than the individual farm data.

The data were split into periods during outbreaks and those occurring when disease was not apparent. Considerable differences were seen in piglet survival between the periods. Moreover, reproductive performance of sows exposed to the PRRS virus was found to be highly heritable. Between 5% and 20% of the differences in survival could be attributed to being heritable, depending on the survival trait in consideration.

Several regions of the genome were found to be associated with piglet survival during PRRS outbreaks. One of these regions overlapped with a previously identified region associated with piglet survival in growing pigs infected with the PRRS virus. Two of the identified regions had been previously linked to disease severity in pigs artificially infected with the PRRS virus, suggesting that similar genomic regions also influence piglet survival in animals infected with the PRRS virus.

The results of this study imply that the devastating effects of PRRS on piglet survival during pregnancy and shortly after birth can be mitigated by genomic selection.

Chapter 1. Introduction and Literature Review

1.01 Porcine Reproductive and Respiratory Syndrome

Porcine Reproductive and Respiratory Syndrome (PRRS) is the most important infectious disease challenge for industry worldwide. As the name suggests the disease manifests itself through effects on both the respiratory and reproductive systems. Respiratory problems mainly affect young pigs resulting in inappetence, lethargy and in some cases death, the virus is the most commonly isolated in cases of the polymicrobial diseases; porcine respiratory disease complex (Terpstra *et al.,* 1991; Brogden & Guthmiller, 2002). The reproductive problems affect pregnant sows and their developing foetuses resulting in reproductive losses, mummification of piglets *in utero*, increased stillbirths and abortions (Zimmerman *et al.,* 2003). In some cases a blue discoloration forming on the ears, vulva or hind (Nodelijk, 2002; Wensvoort *et al.,* 1991; Zimmerman *et al.,* 2003). It is the most economically significant disease affecting pigs, estimated to cost the industry €1.5 billion annually in Europe (de Paz, 2015) and $664 million annually in the US (Holtkamp *et al.,* 2013). PRRS is reputed to be the costliest viral disease in the world (Dekkers *et al.,* 2017), however, when direct and indirect costs are considered the HIV pandemic is likely to far exceed this (Hutchinson *et al.,* 2006). Whereas for human diseases in many countries governments bear much of these costs, for PRRS, the losses are borne by the pig production industry.

Originally identified in the US in 1987 (Rossow, 1998), the aetiological agent of PRRS is the Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) (Terpstra *et al.,* 1991; Wensvoort *et al.,* 1991). The clinical signs provide the basis of some of the synonyms used for the disease including: blue ear pig disease, mystery swine disease; Swine Infertility and Respiratory Syndrome (SIRS); Porcine Epidemic Abortion and Respiratory Syndrome (PEARS) and Lelystad Virus (LV) (Zimmerman *et al.,* 2003). PRRS and PRRSV are now the generally accepted terms for the disease and the virus respectively. The name Lelystad

Virus has been retained to refer to a group of strains related to the original virus identified by Wensvoort *et al.,* (1991).

Originally considered a single species of virus, Kuhn *et al.,* (2016) proposed recognition of two distinct PRRSV species, *Porcine reproductive and respiratory virus 1* and *Porcine reproductive and respiratory virus 2*, following a phylogenetic analysis including eleven, previously undescribed, related, simian viruses. Following ratification by the International Committee on Taxonomy of Viruses (ICTV), the two separate species were recognised as separate members of the family *Arteriviridea* within the order *Nidovirales* (Adams *et al.,* 2016). Subsequently, updates in 2017 created the genus *Porartevirus* which grouped *Lactate dehydrogenase-elevating virus*, *Porcine reproductive and respiratory virus 1*, *Porcine reproductive and respiratory virus 2* and *Rat arterivirus 1* (Adams *et al.,* 2017). At the same time two similar species, previously considered within the same taxon as PRRSV; *equine arteritis virus* and *simian haemorrhagic fever virus, were* moved to two new genera *Equartevirus* and *Simartevirus* respectively (Adams *et al.,* 2017). Viral taxonomy is a dynamic subject and is continuously revised, to incorporate updated *in silico* taxonomic information, based on whole-genome sequence data into traditional polythetic taxonomic classification based on shared characteristics and ecological niche (Van Regenmortel, 1989; Zhang *et al.,* 2012; Kuhn *et al.,* 2016). In recent literature the term PRRSV is still used to refer broadly to the viruses of both species, with a distinction provided as necessary (e.g. Rahe & Murtaugh, (2017)), a convention that will be used throughout this thesis.

Earlier phylogenetic analyses of *Arteriviridea* including lactate dehydrogenase-elevating virus (LDV) and equine arteritis virus (EAV) suggested a shared common ancestry between the PRRS viruses and LDV (Mardassi *et al.,* 1995; Meng *et al.,* 1995; Murtaugh *et al.,* 1995; Nelsen *et al.,* 1999) which is reflected in the new classification (Kuhn *et al.,* 2016; Adams *et al.,* 2016, 2017). Whilst the distinction between the two species has always been recognised, the viruses were originally defined as two distinct genotypes PRRSV-1 (Genotype 1 or

European) and PRRSV-2 (Genotype 2 or North American) (Keffaber, 1989; Wensvoort *et al.,* 1991; Nelsen *et al.,* 1999; Kuhn *et al.,* 2016; Adams *et al.,* 2016). The two species are genetically distinct with a nucleotide identity between the two of only 55-70% (Allende *et al.,* 1999; Nelsen *et al.,* 1999; Lunney *et al.,* 2016).

PRRSV are enveloped viruses with a single stranded positive sense RNA genome of approximately 15 kb (Conzelmann *et al.,* 1993; Spilman *et al.,* 2009). Originally isolated from porcine alveolar macrophages (PAMs) (Wensvoort *et al.,* 1991), these viruses primarily replicate within cells of macrophage and monocyte lineage with a fairly narrow tropism (Wensvoort *et al.,* 1991; Voicu *et al.,* 1994; Duan *et al.,* 1997b, 1997a).

Methods of horizontal transmission of PRRSV been identified including direct contact (Yoon *et al.,* 1993; Albina *et al.,* 1994; Wills *et al.,* 1997) with animal aggression reported as a risk factor (Bierk *et al.,* 2001) or sexually transmitted via infected semen (Yaeger *et al.,* 1993; Gradil *et al.,* 1996; Prieto *et al.,* 1997b). Indirect transmission has also been shown through; fomites (Dee *et al.,* 2002, 2003; Pitkin *et al.,* 2009); insect mechanical vectors (Otake *et al.,* 2002c, 2003); aerial transmission (Kristensen *et al.,* 2002; Otake *et al.,* 2002a); and percutaneously via medical equipment such as hypodermic needles (Otake *et al.,* 2002b). Several field studies have proposed aerial transmission as a likely explanation for outbreaks (Komijn *et al.,* 1991; Mortensen & Madsen, 1992; Vannier, 1993; Mortensen *et al.,* 2002; Zhuang *et al.,* 2002), however such analyses may be unable to resolve confounding of aerial transmission with insect mechanical vectors. PRRSV may be detected at distances of up to 9.1 km from infected farms, with infectiousness of the collected particles confirmed using permissive MARC cells (Dee *et al.,* 2009; Otake *et al.,* 2010), similar methods have been used to show that the infectiousness of airborne particle can relate to particle size (Alonso *et al.,* 2015). Attempts at experimental aerial infection have had mixed results limited to short distances though this demonstrates a proof of concept for transmission of the virus by airborne means (Torremorell *et al.,* 1997; Wills *et al.,* 1997; Kristensen *et al.,* 2002; Otake *et*

al., 2002a). The virus is sensitive to environmental conditions such as drying, pH ranges outside 5-7 and temperatures above 37ºC (Benfield *et al.,* 1992; Bloemraad *et al.,* 1994). Stability at lower temperatures may explain the higher incidence of outbreaks in the autumn and winter months in the US (Tousignant *et al.,* 2015).

Following viral entry into the host, the initial target is the PAMs of the pulmonary mucosa where entry occurs through receptor-mediated endocytosis (Duan *et al.,* 1998). Models have been proposed for viral adsorption, internalisation and the release of the viral genome into the cell for subsequent replication (Van Breedam *et al.,* 2010). Studies having shown the significance of Heparan sulphate (Jusa *et al.,* 1997), Sialoadhesin (Duan *et al.,* 1998, Vanderheijden *et al.,* 2003) and CD163 (Calvert *et al.,* 2007) as receptors for viral entry to the cell, for successful and prolific replication.

Once inside the host the virus employs a number of mechanisms by which it evades host immunity, able to affect both the innate and adaptive immune response (reviewed in Loving *et al.,* (2015)). Immunosuppression of the host occurs via immunomodulation (Van Reeth *et al.,* 1999; Albina *et al.,* 1998; Darwich *et al.,* 2010; Renukaradhya *et al.,* 2010; Yoo *et al.,* 2010) and inhibition of PAM function (De Baere *et al.,* 2012), leading to increased coinfections and high rates of mortality (Thanawongnuwech *et al.,* 2000; Murtaugh *et al.,* 2002; Renukaradhya *et al.,* 2010). The immune response to PRRSV infection is a complicated, if fascinating subject, beyond the scope of this introduction to the PRRSV host-pathogen interaction, more comprehensive reviews of the subject can be found in Murtaugh & Genzow (2011); Amadori & Razzuoli (2014) and Lunney *et al.,* (2016).

Whilst there is no evidence of differences between PRRSV-1 and PRRSV-2 in the pathogenesis of disease in the boar (Han *et al.,* 2013), differences in virulence are observed in the sow (Halbur *et al.,* 1996), though both species have been shown to manifest respiratory disease and reproductive failure (Pol *et al.,* 1991; Wensvoort *et al.,* 1991; Collins *et al.,* 1992; Halbur *et al.,* 1993; Mengeling *et al.,* 1994). It is also observed that differences

in the type of reproductive failure (abortion/mummified/stillborn) is dependent on the age of sow and the stage of gestation (Zimmerman *et al.,* 2003).

In the reproducing sow, the mechanisms that lead to reproductive failure have been explored but are not fully understood. Experimental infection during early embryonic development (at gestation day 7, 14 and 21) detected virus in none of the dead (0/18) and very few of the live (4/131) foetuses, recovered by necropsy at 21 days post infection (dpi); despite significant mortality in the infected group as compared to the control (Prieto *et al.,* 1996). Considerable transplacental transmission was observed following experimental infection at gestation day 90 in all inoculated sows in Karniychuk *et al.,* (2011), however histopathological lesions indicative of PRRS pathology were absent in organs from necropsied foetuses. Using TUNEL (a fluorochromic staining technique to identify DNA damage associated with apoptosis) evidence was seen of significantly greater levels of apoptosis in the implantation sites at the foetus/maternal interface in infected sows. It is proposed in Karniychuk *et al.,* (2011, 2012) and Karniychuk & Nauwynck, (2013) that foetal death may not occur as a result of active infection in the foetus but as a result of apoptosis in the implantation sites and detachment from the uterine wall. Transplacental transmission from sow to foetus following infection of the sow at gestation day 85 was shown in (Ladinig *et al.,* 2014d, 2015a, 2015b). Viral infection of the foetus and levels of virus at the maternal interface were significant predictors of foetal death (Ladinig *et al.,* 2015a). A further analysis of dead and aborted foetuses from PRRSV outbreaks on farms in Thailand suggested that the risk of virus being detected in the foetus differed with the age of the foetus at death (the size of mummified piglets used to infer the age at which they died), though this effect was not significant in the 89 piglets analysed (Olanratmanee *et al.,* 2015). PRRSV has been also been isolated from ovarian tissue and ovarian follicles though this has not been associated with decreased levels of reproductive success (Prieto *et al.,* 1997a, 1997b; Sur *et al.,* 2001). The correlation between risk of foetal death and detectable virus in the foetus has led some to challenge the

idea that foetal death occurs as a result of focal detachment (Lunney *et al.,* 2016). It is possible that foetal death occurs as a result of direct infection of the foetus though other explanations for the observations could be; different mechanisms causing death at different stages of development or, more likely, more developed foetuses later in gestation, surviving just long enough following endometrial infection to permit infection and viral replication within the foetus.

Whilst the mechanisms behind PRRSV related reproductive failure may remain unclear, studies have demonstrated the usefulness of measures of reproductive failure as an indicator for disease (Nodelijk *et al.,* 2003; Schukken *et al.,* 1992).

1.02 Approaches to Tackling PRRS

The pig industry has a number of options to tackle PRRS depending on the type of farm being managed, biosecurity and health status, the frequency (or risk) of PRRSV outbreaks, the age group and flows of pigs and/or semen onto the farm and the aims of the strategy (eradication or management). Several methods are commonly used to tackle a PRRSV outbreak: remove all animals from the farm (depopulation), disinfecting all the facilities and repopulating with high health animals (depop/repop); multi-site isowean facilities (a method of isolating the weaning animals on separate sites) (Whiting & Pasma, 2008); implementing a selective approach to the removal of animals based on age and direct PRRSV testing, medicating against bacterial infection[1] and replacement of stock with high health animals (also known as the Plomgaard Method) and vaccination strategies with either conventional commercial or autogenous vaccines (Dee *et al.,* 1993; Harris, 2000; Linhares *et al.,* 2015; Murtaugh & Genzow, 2011). All these methods require herd closure or acceptable losses and

[1] Whilst treatment with antibiotics will have no direct effect on viral infection, including PRRSV infection, it may reduce the impact of opportunistic co-infections with bacterial pathogens which are often the causes of losses in PRRS outbreaks.

are therefore costly to implement and are not able to confer sustained protection to subsequent PRRSV outbreaks.

Attempts to develop a comprehensive vaccination strategy to combat PRRS have had limited success, with several issues affecting the efficacy of vaccine only control strategies. Vaccine-mediated eradication approaches rely on stimulating the production of Neutralising Antibodies (NAb) which confer protection against heterologous strains (Burton, 2002). Currently available vaccines only confer limited cross protective immunity against heterologous viral challenge (Zuckermann et al., 2007; Dwivedi et al., 2011; Vu et al., 2011). The inability of animals to produce cross protective immunity to heterologous PRRSV infection is thought to occur due to the diverse pathology of the virus including suppression of innate immunity, shielding of NAb targets and rapid evolution leading to extensive genetic variation of quasispecies (Meng, 2000; Butler *et al.,* 2014). A further risk or problem arises from the use of modified live vaccines as the vaccine strain may revert to virulence and result in a PRRSV outbreak (Storgaard *et al.,* 2001). One pragmatic autogenous vaccination strategy used in some countries involves collecting serum from adult animals currently on farm and using a pool of such sera to vaccinate replacement females imported to the farm.(Zimmerman *et al.,* 2003)

Given the costs of managing the condition, inconsistencies in instigating a protective immune response to PRRSV and absence of an eradication strategy, farm management of the PRRS risk requires a multifaceted approach incorporating a range of approaches (Lunney and Chen, 2010). As discussed below, there is evidence for genetic variation in response to PRRSV infection. Thus, it has been argued by the late Professor Stephen Bishop and others that there may be opportunities to identify host genotypes that are less susceptible to PRRSV or its effects. as a contribution to controlling PRRS (MacKenzie & Bishop, 1999; Davies *et al.,* 2009; Lunney & Chen, 2010). More recently, reverse genetic approaches, in which gene

editing of genes encoding molecules involved in viral entry and genome release, have also been explored as means to control PRRS (Whitworth & Prather, 2017)

1.03 Genetic Improvement of Livestock

Selecting for traits of interest in animals and plants (whether actively or passively) is a striking outcome of human interaction with other species this has resulted in the domestication of a range of animals; including the domestic pig, *Sus scrofa domesticus* (Diamond, 2002). Following the development of agriculture and a greater understanding of the patterns of inheritance specific tools have been developed to improve the efficiency of the selection process. Fisher, 1918 first coined the term "Variance of the normal population" highlighted the desirability of identifying the "constituent causes fractions or percentages or percentages of the total variance" and related this to patterns of Mendelian inheritance. The most widely used tool for assessing an individual's merit with regards to trait variance, is Best Linear Unbiased Prediction (BLUP) (Henderson, 1949, 1950, 1973). Traditionally this technique brings together information about the traits of interest (the phenotype) with information about how animals are related; to calculate the genetic merit or estimated breeding value (EBV) of the individuals with regards to the trait(s) under consideration. Using BLUP EBVs can be estimated for animals without direct phenotypic measures based solely on their relationship to phenotyped individuals (Henderson, 1977). Several models can be used in establishing breeding values, however I will focus on the animal model used throughout this thesis. In the animal model, traditionally, a pedigree is used to calculate a matrix representing two times the coefficient of co-ancestry for all animals in the pedigree also called the additive genetic (or numerator) relationship (A) matrix. This A matrix, fitted as a random effect incidence matrix in a linear mixed model, estimates variance in the breeding values or additive genetic variance (σ^2_A) with a remaining residual error or environmental variance (σ^2_E) (Lynch & Walsh, 1998). The sum of the two variances (σ^2_A + σ^2_E) gives the phenotypic variance (σ^2_P) and the ratio of the additive genetic variance to the

phenotypic variance provides an estimate of narrow sense heritability (h^2) (Falconer & Mackay, 1996). Heritability is important in both selective breeding and evolution as this proportion of variation is controlled by additive genetic factors and thus is the component which may be exploited by both natural and artificial selection processes. The change in trait mean in one generation (ΔZ) is described in the breeder's equation as the product of narrow sense heritability (h^2) and the selection differential (S) (Lush, 1937). The selection differential, or fitness advantage conferred by the trait, can be controlled in selective breeding by selecting the animals contributing to subsequent generations, i.e. parents.

The additive relationship matrix is based on the concept of identity-by-descent (IBD), the probability that alleles across the genome are inherited from a common ancestor (Wright, 1921, 1922). More recently tools have been devised to incorporate molecular genotypes as assayed using SNP (single nucleotide polymorphism) chips and use identity-by-state (IBS) to calculate a genomic relationship matrix (GRM or G matrix) which can be used in place of, or alongside, the A matrix (VanRaden, 2007) in genomic best linear unbiased prediction or GBLUP.

Further tools have been developed to localise these heritable effects to regions of the genome. These tools use the marker genotypes (e.g. SNP genotypes) as an indication on the effects associated with that region based on linkage. Based on the work of Morgan (1911) linkage reconciles observations of coinheritance with the much earlier theory of the segregation and random assortment of genes (Mendel, 1865). The degree of linkage in a region can be assessed by measures of linkage disequilibrium (LD). LD describes the frequency with which alleles at separate loci on the genome are non-randomly coinherited (Lewontin & Kojima, 1960). Methods of quantifying LD vary though one of the most commonly used methods in population based studies is the squared coefficient of correlation (r^2) ranging from 0 (no correlation) to 1 (perfect correlation). This method provides a standard means of comparing LD between loci (Weir, 1979).

These core concepts in genetics have been developed into highly successful tools for probing the genome to identify areas associated with variance in given phenotypes. One such method, genome wide association (GWA) analysis has become an increasingly popular method by which the relationship between phenotype variance and variation in genotypes can be correlated. This tool can be used to identify Quantitative Trait Loci (QTL), specific variation of the genome which explain variance in the phenotype of interest. The variant in the genome with an effect, cannot be said to be necessarily causal, but by virtue of linkage can provide a useful indication of areas for further study or selection. The presence of linkage allows large sections of the genome to be scanned for effects without the need for all potential polymorphisms in the genome to be genotyped (Zondervan, 2011).

The use of genomic tools to asses and ascribe fitness has given rise to the development of genetic methods for the selection of animals in breeding programmes, such as marker-assisted selection (MAS) and genomic selection (GS). In MAS, the selection candidates are screened for the genetic variant(s) with effects identified using association methods. This can be used with breeding values for other traits of interest, either in tandem or using an index ascribing a value to the genetic variant(s) of interest (Dekkers, 2004). In genomic selection animals with both phenotype records and genotype data (the training population, reference population or predictor population) are used to calculate genomic estimated breeding values (GEBVs) for selection candidates (the validation population or predictant population) based solely on their genotypes. In the training population marker effects are calculated based on the reference population using association methods; in the validation population the individual's GEBVs are estimated based on the sum of the effects of their markers (Meuwissen et al., 2001; Hayes et al., 2009b). This process is called genomic prediction and relies on LD between the markers and mutations responsible for trait variance (Goddard & Hayes, 2007).

Genomic selection, in addition to improved accuracy offers additional advantages. Given it is not necessary to have direct measures of phenotype in selection candidates, the choice of animals as future parents can be made earlier, genetic merit can be estimated for both sexes for traits only expressed in one (i.e. milk production) and traits can be evaluated in live animals that could only be assessed *post-mortem* or later in life (certain carcass traits or degenerative conditions) (Hayes & Goddard, 2010). This approach would also be the preferred method of estimating genetic merit in disease resistance given it is undesirable to directly challenge selection candidates for introduction into high health genetic nucleus herds.

Successes has also been made using gene editing technology to modify pigs for resistance to PRRSV (Prather *et al.,* 2013; Whitworth *et al.,* 2015; Burkard *et al.,* 2017; Wells *et al.,* 2017; Whitworth & Prather, 2017). This approach has been used to try and block infection at the viral entry stage by deleting or modifying the genes encoding the molecules involved in viral binding, entry and genome release. The initial target was sialoadhesin (CD169), but pigs lacking CD169 remained susceptible to PRRSV infection (Prather *et al.,* 2013). More recently, the focus has switched to the CD163 gene on chromosome 5 (SSC5) that encodes a p155 or haptoglobin scavenger receptor which has been shown to be involved in facilitating the release of the PRRSV genome. The results from three studies have been published. In the first study, the CD163 gene was disrupted or deleted using the CRISPR/Cas9 system (Whitworth et al 2015). The resulting edited pigs were resistant to infection with PRRSV. The other two studies have focused on exon 7 that encodes cysteine-rich domain 5 (SRCR5) one of nine extracellular scavenger domains (Burkard *et al.,* 2017; Wells *et al.,* 2017). While CD163 has been implicated in a broad range of immunological functions, the SRCR5 domain has not been implicated in any of the functions associated with the receptor (Van Gorp *et al.,* 2010; Burkard *et al.,* 2017). Cells from pigs in which exon 7 has been excised using CRISPR/Cas9 tools have been shown to be resistant to infection with both PRRSV-1

and PRRSV-2 and to retain some of CD163's potential important functions (Burkard *et al.,* 2017). Whilst there is no biological function currently associated with SRCR5 further work will be needed to characterise the effects of this deletion *in-vivo* under field conditions. In contrast, replacing the pig CD163 exon 7 with the homologous sequence from the human CD163L1 gene did not result in PRRSV resistance (Van Gorp *et al.,* 2010; Wells *et al.,* 2017).

While the success in gene editing offers a huge opportunity from a commercial perspective, the regulation of edited animals is unclear and the public debate on their acceptance is still ongoing (Hoyos-Flight *et al.,* 2017). Whether existing regulatory definitions of genetically modified organisms (GMOs) apply to organisms edited using CRISPR/Cas9 technology is still unknown (Hartley *et al.,* 2016). In the UK, given the lack of regulatory guidelines on the use of gene edited animals; political debate is likely to delay bringing any viable product to market. This issue will vary between jurisdictions.

1.04 Selecting for Health Traits in Response to Disease

The traits to use for the genetic improvement of the health of livestock in response to disease have been widely debated. Typically, two terms are used for describing two conceptually different manners in which the effects of a specific pathogen are mitigated; that of resistance and tolerance. Resistance is generally accepted as the ability of the host organism to limit pathogen burden or indeed infection. Whereas tolerance is performance despite pathogen burden (Albers *et al.,* 1987) or, mathematically; the slope of the regression of performance on disease burden (Simms & Triplett, 1994; Bishop, 2012). The wider term resilience maybe used to describe performance both in the presence or absence of disease challenge. Models have shown that selection for resistance with respect to PRRSV should reduce the impact and frequency of PRRSV infections (MacKenzie, 1999; Bishop & MacKenzie, 2003). Though whether to select for resistance or tolerance traits is a much-debated issue. Resistance may offer advantages, in that potentially only a proportion of the animals need to

exhibit total resistance in order to confer herd immunity to the group (MacKenzie & Bishop, 1999). Whilst this reduction in pathogen reproduction reduces the opportunity for the pathogen to evolve to overcome the resistance mechanism; the emergence of pathogen genotypes that could circumvent the host resistance mechanism could still occur in other (non-resistant) populations. The functional basis of resistance is of considerable importance to the risk of the pathogen overcoming resistance mechanisms with regards to the opportunities available for such (Bishop & Woolliams, 2014). Selecting for multiple resistance mechanisms could help alleviate any risk (Bishop & MacKenzie, 2003). Outbreak type and prevalence also affect the merits of selecting for tolerance or resistance, endemic diseases or diseases in which the pathogen replicates in reservoirs of other species, will benefit from selection for tolerance given the ubiquitous nature of the pathogen, however the cost benefit ratio will be low if the disease is of a low prevalence (Bishop & MacKenzie, 2003; Bishop & Woolliams, 2014).

Finally a broader, concept or robustness may be applied to improved performance across external stressors (whether pathogenic or environmental) (Knap, 2005). In this wider context of improving performance it may be advantageous in a field setting to consider performance across sources of stress (Archibald *et al.,* 2008).

Genetics of variation of response to disease is increasingly being exploited in breeding programmes. One of the earliest programmes to produce a resistant line commenced in 1987 on Australian Merino sheep for resistance to nematode infection. This selection was conducted using traditional methods on phenotypic mean to 1991, followed by BLUP. Comparing the line selected on the faecal egg count trait (Rylington) to a line where selection was not used a 2.7% annual genetic gain was realised in terms of improvement to mean herd EBV (Karlsson *et al.,* 2006).

Success has also been reported in control of Scrapie in sheep. Amino acid substation variants at three positions in the prion protein (PrP) gene (codon 136, 154 and 171) were strongly

associated with Scrapie susceptibility; identifying both a highly susceptible and a completely resistant variant (Baylis *et al.,* 2004). Widespread programs to limit the susceptible genotypes in the breeding stock were implemented throughout Europe and North America, including the National Scrapie Plan for Great Britain (Dawson *et al.,* 2008). A review of the National Scrapie Plan for Great Britain addressed concerns regarding increases in inbreeding and the effects of the favourable genotype on performance, concluding that selection was feasible and should reduce the impact and frequency of scrapie outbreaks (Dawson *et al.,* 2008). These benefits were demonstrated in a subsequent analysis on the impact of the Dutch breeding programme which selected for resistant rams which found, not only a reduction in prevalence as direct result of selective breeding, but also indirect reductions in the remaining susceptible population (Hagenaars *et al.,* 2010). Compared to other communicable diseases, the pathogenesis of prionic diseases may suggest different underlying host genetic liabilities, however; as a communicable infection similar, epidemiological factors will be applicable.

A further success in controlling disease using selection methods has been reported for Infectious Pancreatic Necrosis (IPN) in salmon. Studies on IPN, a viral disease of Salmonids identified a major single QTL in the Atlantic salmon explaining between 80 and 100% of the observe genetic variation in resistance traits (Houston *et al.,* 2008; Moen *et al.,* 2009). Given the simple nature of the single allele found, it was suitable for use in MAS to select for this resistant allele. Following MAS and widespread use of the selected animals in Norway reductions of 75% were observed in the number of outbreaks in the three years following the introduction of the selected genetic line (Moen *et al.,* 2015).

More recently a new breeding index was released to improve bovine Tuberculosis (bTB) outcomes in cattle. ‘TB Advantage’ was created in a collaboration between the University of Edinburgh, Roslin Institute and Scotland’s Rural College (SRUC). (The Veterinary Record, 2016; Winters, 2016). This index combines pedigree information from bulls with offspring on bTB challenged farms with pedigree selection methods to calculate a genetic index with

regards to bTB susceptibility (Banos *et al.,* 2017; http://www.ed.ac.uk/roslin/research/isp/control-infectious-diseases/genetic-basis-of-host-resistance/breeding-tuberculosis-resistant-cattle, 2017; https://dairy.ahdb.org.uk/technical-information/breeding-genetics/tb-advantage/, 2016).

1.05 Genetic Variation in Response to PRRS

The genetic improvement of animals for PRRSV response relies on the existence of heritable variation of response to PRRSV infection. This has been identified in a range of studies exploring resistance, tolerance and resilience type traits. Understanding the variance of response to PRRSV is important to understanding the variance which can be attributed to underlying genetic differences.

Several *in vivo* challenge experiments have found significant differences between breeds and genetic lines in the variance of clinical signs and pathological response associated with infection with PRRSV (Halbur *et al.,* 1998; Christopher-Hennings *et al.,* 2001; Petry *et al.,* 2005; Vincent *et al.,* 2006; Doeschl-Wilson *et al.,* 2009; Reiner *et al.,* 2010). Petry et al. (2007) reported patterns of differential expression of immunological markers between the highest and lowest respondents by rank. *In vitro* studies have also demonstrated differences between breeds in the susceptibility of immune cell models to PRRSV both in terms of the relative numbers of monocyte derived macrophages infected (Vincent *et al.,* 2005) and replication of the virus in PAMs (Ait-Ali *et al.,* 2007).

Measures of reproductive success/failure is an easily measurable, quantitative trait, measured as part of normal business-as-usual activity on farms. As widespread testing for the presence of PRRSV as part of rigorous biosecurity protocols on farms, reproductive data in the presence and absence of PRRS is available representing a useful source of data on reproductive performance in response to disease.

Reproductive failure in response to PRRSV infection is well characterised both experimentally and under field conditions (Nodelijk *et al.,* 2003). It is concluded by Schukken *et al.,* (1992) that reproductive performance is an informative quantitative measure for the statistical analysis of PRRS. Case control studies showing a significant association with PRRSV seroconversion and aberrant reproductive outcomes (Nodelijk *et al.,* 2003). The same study reported that impacts on herd reproductive performance maybe observed during epidemic type outbreaks, though these are not apparent when the disease is endemic within the herd.

Several studies conducted on reproductive record data are reported in Lewis *et al., (*2009b, 2009a, 2009c). The authors investigated the variance of PRRSV response in terms of reproductive outcomes in a Chinese multiplication unit, the data for which, forms part of the data used in this thesis.

The population from which the field data were obtained experienced two PRRS outbreaks and the data were partitioned using two methods a threshold method and a date method. Using the threshold method, the mean mummified piglets per litter on a rolling basis per day was calculated and the 99% CI of a baseline (PRRS free period) used to partition dates when the trend was above this threshold. Litters farrowing in periods above this threshold were partitioned into a disease phase. A second parallel method, the date method, was also used to define disease phases using the month in which infection was confirmed using ELISA (Lewis *et al.,* 2009b).

Significant differences in reproductive outcomes were demonstrated by sow line and parity. Estimates of least squares means by parity observed a non-systematic parity effect, though parities 1-5 experienced greater losses than parities 6+. Partitioning the animals into Chinese derived lines (with generally larger litters with lower growth rates) and European lines (with generally smaller litters and higher growth rates) suggested any improvements in terms of litter size using Chinese breeds maybe lost under PRRSV challenge (Lewis *et al.,* 2009a).

Heritability was estimated over a range of reproductive traits in data partitioned under both methods (Lewis *et al.,* 2009b). The analysis by Lewis and colleagues found greater contrast between non-disease and disease phase data using the threshold method which consistently demonstrated a higher heritability in disease phase for traits associated with the clinical signs of PRRS (numbers of mummified, dead, and weaned piglets per litter and the number of services required for a successful mating). Using a logistic regression approach heritability was also estimated representing these four quantitative traits as binary traits with a threshold of >1 to distinguish the afflicted and unafflicted condition. A similar difference is observed in the heritability estimates using the binary analysis between disease and non-disease phases, higher disease phase heritability estimates were observed in the binary trait analysis. These results demonstrate that the genetic merit of animals in terms of reproductive performance may be used to mitigate some of the effects associated with PRRSV outbreaks. However, there is a potential source of bias in the inaccurate classification of afflicted and unafflicted animals.

The first published GWA analysis for PRRS disease traits is given in (Lewis *et al.,* 2009c) using a custom 7K SNP chip on the disease phase data. Eleven SNPs were presented as showing genome wide significance throughout the genome. The percentage of the total σ^2_A explained by the SNP variance (σ^2_{SNP}) was between 4.7% and 0.1% of the total baseline estimates. The locations of the four mapped associations presented were SSC1 (98.3 Mbps) SSC4 (92 Mbps), SSC4 (91Mbps) and SSC14 (28.5Mbps) (Lewis, 2008). [Note these genome coordinates refer to an earlier version of the pig reference genome sequence, i.e. prior to the published draft Sscrofa10.2].

The US PRRS host genetics consortium (PHGC) conducted 14 experimental infection trials, see Lunney et al. 2011 for the full experimental protocol. In each trial between 109 and 507 animals from the same or similar genetic backgrounds were experimentally infected with one of two PRRSV strains; NVSL-97-7895 or KS-2006-72109 (characterised in Ladinig et al.

2015). Animals were experimentally infected intranasally and intramuscularly with PRRSV, and animals were weighed and blood samples were collected at regular intervals for a period up to 42-day post-infection (dpi). Blood samples were assayed for viremia using a RT-PCR (reverse transcriptase - polymerase chain reaction) assay for the presence of PRRSV RNA.

An inconsistency was observed in the viremia profiles of some animals (log_{10}(viremia) over time). For some animals after 21 dpi, once viremia had peaked and viral clearance had begun, some animals exhibited a second viremia peak, termed rebound. Using Bayesian inference three statistically different viremia profiles were identified: cleared (unimodal and undetectable at 42 dpi), persistent (transient detectable levels of virus at 42 dpi) and rebound (biphasic, two peaks within the 42-day experimental period) (Islam et al., 2013). These were modelled on the Wood's curve (used to model lactation in dairy herds), for biphasic viremia the Woods curve model was extended to accommodate a second peak. When the extended model improved fit (CI>95%) animals were considered biphasic and the extended model used. Whilst the persistent condition could be predicted based on viremia 0-21 dpi, rebound could not, which could indicate that the rebound condition was triggered by an event unrelated to gross levels of circulating virus (such as viral mutation). Animals exhibiting persistence were found to have higher levels of cross-protectivity to homologous strains of the virus (Islam *et al.,* 2013).

Subsequent sequencing of rebound virus quasispecies found a high degree of variability in open reading frame (ORF) 5, as a result of negative selection (dN/dS ratio). A smaller number of residues showed positive selection. These results suggest that escape mutants are a key component of the generation of quasispecies in the host, that adaptive immunity is likely to play an important role in the diversification of the virus *in vivo*. The small amount of positive selection indicates a potential for the genetics of host adaptive immunity in viral diversification, and as such, control (Chen *et al.,* 2016).

The genetic analysis of the results from PHGC trials 1-3 is presented in Boddicker *et al.,* 2012; with further validation work in trials 1-5 in Boddicker *et al.,* 2014b and using trials 1-9 in Boddicker *et al.,* 2014a with isolate NVSL 97-7895. The two traits explored were viral load (area under the curve of $\log_{10}$(viremia) by time) truncated after 21 days (before rebound is observed) and weight gain up to 42 days. High/moderate levels of heritability were estimated over the 12 trials for viral load and weight gain at 0.44 (s.e. 0.13) and 0.29 (s.e 0.11) showing a strong negative genetic correlation at -0.46 (s.e. 0.20). The h^2 of rebound estimated for trials 1-3 was very low or ~0.

A single 0.5 Mbps region on SSC4 identified in Boddicker *et al.,* (2012) consistently showed a significant association across trials with both weight gain and viral load. Six SNPs within this region show high levels of LD. A single SNP (WUR10000125) was found to fully capture the effects of this region with the SNP variance explaining 15% and 11% of the total genetic variance for viral load and weight gain respectively (Boddicker *et al.,* 2012, 2014b, 2014a). Differential expression analyses showed significantly higher expression of GBP5 in heterozygotes containing the favourable WUR10000125 B allele as compared to the AA genotype at multiple time points over the experimental period. The guanylate binding protein (GBP) gene family is located on SSC4 in the region identified in the GWAS analyses. This region of the draft reference genome sequence (Sscrofa10.2; (Groenen *et al.,* 2012)) is poorly resolved. Sequencing of the GBP5 gene found a premature stop codon associated with the A allele, leading to the production of a truncated and inactive protein (Koltes *et al.,* 2015). It is proposed by Schroyen *et al.,* 2016 that the complete GBP5 protein inhibits viral internalisation and replication and that the incomplete protein (i.e. the form associated with unfavourable WUR10000125 A allele) loses this ability. Macrophage levels of GBP5 having been shown to have a strong inverse correlation with HIV-1 viral levels (Krapp *et al.,* 2016).

The genetic analysis on trails using KS-2006-72109 is presented in Hess *et al.,* 2016. This analysis also incorporated Wood's function parameters to both estimate viral load as a trait

and consider key features of the viral load profile across all 15 trials. From the measures of viremia profiles fitted using the Wood's curve, values of peak viraemia, time to peak viraemia, maximal decay rate and time to maximal decay rate were estimated, enabling a more comprehensive assessment of viraemia over time. Daily weight gain was interpolated using random regression and the effect of the WUR10000125 genotype was also considered. Weight gain and viraemia profiles compared between isolates, suggest that NVSL-97-7895 is more virulent than KS-2006-72109 as indicated by higher peak viraemia, earlier time to peak viraemia and slower growth rate, however virus clearance happened earlier (time to maximal decay rate) and quicker (maximal decay rate) and persisted longer (higher viraemia at 42 dpi) in KS-2006-72109. Estimates of heritability were again generally moderate/high. For viral load, peak viraemia and weight gain, genetic correlations were high between strains, suggesting variance in these traits is cross-protective with regards to strains of differing levels of pathogenicity.

A further study using farm data from reproducing sows during a confirmed PRRS outbreak is reported by Serão *et al.,* 2014. Traits used included reproductive outcomes and PRRSV Ab ELISA sample-to-positive (S/P) ratio ELISA is a colorimetric assay for identification of a test substance, in which an antibody/enzyme conjugate is adhered to a sample and the coloured enzyme product measured using optical absorbance. S/P ratio is the ratio of the mean sample optical absorbance (above a negative control) to a mean positive control optical absorbance (above a negative control) providing an indication of amount of test substance in the original sample. The sample used was blood drawn from animals on a single day during the height of the epidemic, in addition to the litter records of sows. Litters were partitioned according to the rolling trend method in Lewis *et al.,* (2009b) into pre-PRRS phase and PRRS phase and the phases analysed separately. There were 4,702 litters farrowed during the pre-PRRS phase and 525 litters in the PRRS phase. Heritability was estimated using pedigree information and Restricted Maximum Likelihood (REML) used to investigate the

effect of the WUR10000125 genotype on trait variance. GWA analysis was conducted using the Bayes B method, fitting a mean preadjusted (for fixed effects) residual by animal used to account for multiple records. Associations were reported as % variance explained by the markers in 1Mbp regions, compared to variance explained by all markers. Regions explaining >5% of the total marker genetic variance, within 2 Mbp, were combined and the analysis rerun to consider the wider area in which the QTL could lie. Candidate SNPs were then selected on the basis of the percentage of the regional genetic variance that the individual SNPs explained. Heritability estimates using the pedigree for reproductive traits were similar to those reported in the literature for pre-PRRS phase, low for loss traits and low/moderate (0.08) for number born alive. Estimates varied in disease phase from low (0.01 s.e. 0.01 for farrowing to weaning mortality) to low/moderate (0.12 for number born dead) for reproductive traits. Estimates were moderate/high for ELISA S/P (0.45 s.e. 0.13). Slightly lower estimates of heritability are reported in disease phase than the pedigree based estimates presented in Lewis *et al.,* 2009b in an independent dataset. Genetic correlations were high between reproductive traits but low between S/P ratio and the reproductive traits. No effect was indicated in the PRRS phase for the WUR1000125 genotype, though several regions in association with traits were presented. In total, 7 discrete regional associations are presented. One on SSC1 at ~32 Mbps explained 11% of Stillborn total marker variance and 0.81% of total marker variance in farrowing mortality percentage. A further 6 regions explained between 0.67% and 25.15% of ELISA S/P ratio trait variance. A region on SSC7 between 24–30 Mbp explained the largest amount of total marker variance and contained the major histocompatibility complex II gene cluster. A further region on SSC7 downstream (128–129 Mbp) explaining 15.73% of the total marker variance.

Further characterisation and validation of the two SSC 7 regions explaining the most total marker variance in Serão *et al.,* (2014) was conducted in Serão *et al.,* (2016). In this analysis between 204 and 368 high health PRRSV naïve gilts from 7 different genetic sources were

monitored before, during and after acclimation on 23 commercial farms with high levels of disease challenge (gilt acclimation data). Methods of exposure to PRRSV varied according to the destination in terms of the use of vaccination strategies (used in 18/23 of the farms) or natural infection. These data were used in conjunction with those presented in Serão *et al.,* (2014) (challenge data). The animals were genotyped and ELISA S/P ratio at the end of acclimation obtained. This study used the same GWAS methodology for the identification of regions with a genomic selection cross validation conducted, predicting the breeding value of animals from each genetic source (the validation dataset) using animals from the remaining six (the training dataset). Repeating the process such that each source was used as the validation dataset. Predictions were always made on contemporary groups where 100% of animals show strong ELISA results, though contemporary groups with a lower proportion were explored for their accuracy as the training dataset. Heritability estimates were lower when estimated on datasets from contemporary groups showing a lower proportion of strong sample to S/P. The highest estimate of heritability was seen using the genotype information with challenge data, which was slightly higher than that previously reported using the pedigree at 0.54. Whilst lower estimates of heritability are to be expected with reduced pathogen exposure (Bishop & Woolliams, 2010), given the different methods by which the sows were exposed, prevalence cannot be inferred from the frequency of strong ELISA responses. Low to moderate cross validation accuracies were reported across the analyses. The highest accuracy reported was when the SNPs in the training and validation method were limited to the 2 QTL regions on SSC7, conversely the lowest accuracies were reported when using the remaining genotypes (excluding QTL regions). SNPs in the MHC QTL (accuracy 0.39 for outbreak data and 0.31 for gilt acclimation data) performed better than the region downstream (0.34 for outbreak data and 0.25 for gilt acclimation data). A large proportion of the genetic variance in PRRSV antibody production was captured by these two regions. Whilst differences were seen between the gilt acclimation study and the farm outbreak data there were some notable differences between the conditions of exposure

(natural/vaccinated), variation in the time of exposure/vaccination and time the ELISA sample was taken (26 to 103 days). Despite the MHC region explaining between 20% to 25% of the total marker variance in this analysis, its suitability for improvement of the host to disease outcomes is questionable. Low genetic correlations were seen between reproductive trait and ELISA S/P in the original study suggesting that higher immune response is not conferring protection in terms of reproductive outcomes. Some studies have argued that diversity in the MHC genes assists in the range of pathogens recognised by the immune system (Sommer, 2005) with evidence of overdominant selection for MHC II diversity (Hughes & Nei, 1989). Selection on the MHC II gene would reduce genetic diversity and may therefore effect the repertoire of immune gene variants to identify other pathogens. While for some avian species reduced variation in MHC genes does not appear to have affected fitness (Gangoso *et al.,* 2012) understanding of the protective role of MHC diversity across pathogens is not complete (Ellis & Hammond, 2014).

Given the snapshot of antibody response taken at one day, and the dynamic nature of disease outbreaks it is possible that some variance in antibody response could arise due to variation in the time each sow was infected. Several studies (reviewed in Adamo, 2004) have suggested that caution should be exercised when using antibody titre as an indicator of disease resistance, tolerance or resilience.

A direct challenge experiment on reproducing sows was also conducted in (Ladinig *et al.,* 2014d). One hundred and thirty-three high health purebred Landrace sows were selected, oestrus was stimulated, synchronised and the sows were inseminated, (re-inseminated as necessary) and conception confirmed. Sows were assigned to an infection or control group, housed separately and infected intranasally and intramuscularly with PRRSV isolate NVSL 97–7895. Daily observations were made, rectal temperature taken and whole blood samples collected. At 21 dpi (~106 gestation days) the animals were euthanised and necropsied. The entire reproductive tract was removed, plus additional tissue samples and the position of

each foetuses and its preservation status documented. Five preservation statuses were used: viable (alive, normal appearance and colour), meconium stained (alive, tan staining, thick brown amniotic fluid), decomposed (dead, no blood in umbilical, generally not discoloured), autolysed (dead, no blood in umbilical, generally discoloured) and mummified (<20cm, dehydrated). Given mummified piglets die <35 days of gestation (Christianson, 1992), prior to infection in this study, mummified piglets were removed from the analysis. Less than 1% of meconium stained foetuses were found in the control group compared to 9% in the infected group, it was therefore considered a PRRS associated pathology and probable that these animals would either not survive to term or die shortly after farrowing. From the phenotypic and immunological measures several correlations were found: a significant association was observed between birthweight of piglets and the birthweight of parent sow, though no significant differences were observed in the levels of virus, cytokines or in the clinical signs used in the study between low birthweight and high birthweight piglets (Ladinig *et al.,* 2014a); significant differences were observed in leukocyte profiles (Ladinig *et al.,* 2014b) and cytokine profiles (Ladinig *et al.,* 2014c) correlated with both viral load and foetal mortality.

A genome wide association study conducted on foetal characteristics based on these data is presented in Yang *et al.,* (2016). Two binary traits were analysed; foetal death (dead at the time of necropsy) and foetal viability (categorised as viable at the time of necropsy). Two continuous traits were also analysed: viral load in the thymus and viral load in the endometrium. GWA analysis was conducted using least absolute shrinkage and selection operator (LASSO) with permutation analysis (×1000) used to calculate empirical experiment wide p values. No association was identified in the region of the major SSC4 QTL identified in Boddicker *et al.,* (2012) nor in the region on SSC7 explaining a high proportion of the total variance in ELISA S/P ratio identified in Serão *et al.,* (2014). Comparing the locations of the QTL identified in the animal QTL database (Hu *et al.,* 2005) cross over was found

with previously reported loci for disease resistance/susceptibility traits, cytokine and leukocyte profile.

1.06 Aims of this Research

With respect to deriving a robust approach to genetic improvement of the domestic pig to PRRSV infection there are a number of questions which remain unanswered with respect to the genetics of host response. Whilst a major QTL for PRRSV resistance has been identified on SSC4; studies using reproducing sows have failed to identify any effect in this region on reproductive losses. Given that piglet mortality reflects the highest component of financial exposure faced by the pig industry (Holtkamp *et al.,* 2013) it is unclear as to whether genetic variation in this region could mitigate substantial amounts of the economic losses associated with PRRS.

Whilst a number of groups have investigated the distribution of PRRS related traits throughout the genome, the results of these findings have been inconsistent with relatively little cross over in the position of the loci identified. Whilst there are differences in the traits used in these analyses (viral load, weight gain and reproductive traits under field and experimental conditions) it may be expected that improved resistance and/or tolerance would translate to reductions in mortality.

The aims of my research as reported in this thesis are to explore the reproductive response of pigs to PPRS outbreaks in field data. These aims will be met using the variance in reproductive traits associated with PRRSV outbreaks, from data collected from two farms experiencing PRRS. In establishing the factors which contribute to reproductive success or failure suitable models will be established for the estimation of heritable variation in the reproductive traits, and non-genetic effects considered. The heritable variation can then be exploited in conjunction with association study methods, to identify regions of the genome to

which the heritable variation can be attributed. Single SNP analyses will be explored in addition to methods to explore the regional effects.

Whilst most of the phenotypic data used in this study have been analysed previously (Lewis *et al.,* 2009a, 2009b, 2009c), the power of these data for genetic analyses have been significantly enhanced by the addition of SNP genotypes generated using the Illumina PorcineSNP60 chip (Ramos *et al.,* 2009; Illumina, 2015) for animals in the original population. In addition, an independent dataset (farm 2) is analysed. It is hypothesised that the increased density SNP chip will enhance the ability to detect QTL by virtue of increased coverage. It is possible that there may be improved power to estimate heritability and additive genetic variance by combining the two farms' data. That this increased power and increased coverage may enable the discovery of new regions of interest in addition to providing an increased dataset with which to consider previously reported QTL.

In summary, this thesis will explore the genetic architecture underlying reproductive outcomes to PRRSV infection:

- To consider the factors affecting non-genetic effects contributing to reproductive trait variance in PRRSV challenged data, such that models may be devised to explore genetic effects.
- To quantify additive genetic variance and heritability, using pedigree information (where available) and 60K SNP genotypes.
- To utilise single SNP association methods to explore the genetic architecture of reproductive trait variance across the pig genome
- To use regional association methods, to look at the reproductive effects in wider genomic windows

Chapter 2. Impact of PRRSV Outbreaks on Reproductive Performance of Sows

2.01 Introduction

As outlined in the previous chapter, a number of studies have investigated the impact of PRRSV on reproductive success in sows. PRRSV has been identified as a suitable target for genetic improvement of the host (Bishop & MacKenzie, 2003; Davies *et al.,* 2009) and research has focussed on the heritable variation of the response of the host to infections with the PRRS virus. A prerequisite for quantifying heritable variation of performance traits, is an understanding of systematic, non-genetic factors that underpin the variation in the observed reproductive performance of sows. This is especially true using inherently noisy data from farm settings, during the dynamic conditions that epidemics present (Bishop *et al.,* 2012).

Variation in reproductive success is reviewed extensively in the literature. Non-pathogenic factors include parity, breed, temperature, permanent environmental effect, nutrition and various toxins. (Christianson, 1992; LeFebvre, 2015). Variation has been explored in reproductive traits in direct response to PRRSV (Lewis *et al.,* 2007, 2009a; Rashidi *et al.,* 2014; Serão *et al.,* 2014) identifying non-genetic effects relevant to the cohorts studied.

The data available to this study is field data collected from two commercial farms, which had experienced PRRSV outbreaks with data available during and outwith PRRSV outbreaks. Previous analyses of the same data on one of these farms found that reproductive success was significantly reduced during the PRRSV outbreak phases and that parity and sow line had a significant effect on reproductive performance traits for sows in both epidemic and non-epidemic phases (Lewis *et al.,* 2009a, 2009b). Although these analyses distinguished between epidemic and non-epidemic phases, they did not account for differences within or between separate outbreaks, which is also expected to have a strong effect on performance and its underlying heritable variation (Bishop & Woolliams, 2010).

The challenges posed in the use of field data for genetic analyses are discussed in (Bishop & Woolliams, 2010), which identified three main sources of uncertainties that affect variance component estimates.

1) The different levels of exposure faced by individuals within the population

2) Inaccuracies in the classification of animals as diseased or healthy

3) The dynamic nature of epidemics within populations

This chapter forms the prerequisite of the genetic analyses carried out in the subsequent chapters. It starts with a detailed description of the datasets on which all studies in this thesis are based. Using a modified version of the threshold/threshold method previously reported (Lewis *et al.,* 2009b), partitioned datasets are created to distinguish between epidemic and non-epidemic phases in both farms; to enable the investigation of the effects of PRRSV outbreaks on the reproductive performance of sows. This is explored using linear mixed models, which also provide estimates of non-genetic fixed and random effects affecting reproductive failure during the PRRSV epidemics. Finally, different ways to account for differences in exposure in the statistical models are explored, and their impact on reproductive performance is assessed. The models created in this chapter are then carried forward in conjunction with population and SNP data to explore genetic parameters.

2.02 Materials and Methods

2.02.1 Qualitative and Quantitative Description of the Data

Data was available for 2 farms provided by GENUS / PIC, termed from here on, farm 1 and farm 2. During the 5- and 2-year periods for which data was available, both farms experienced a minimum of one PRRSV outbreak confirmed by a commercial enzyme linked immunosorbent assay (ELISA) test with a sensitivity of 97.4% and specificity of 99.6% at

herd level (Idexx Herdchek, 2003). These data were cleansed and standardised for subsequent, data analyses following the procedure outlined below.

Farm 1

Farm 1 was a high health, Chinese multiplication unit subject to monthly ELISA testing for PRRSV antibodies. It contained 14 sow lines representing animals from separate sources, breeds and crosses. Purebred lines included, Large White, Landrace, Pietrain composite, Duroc and Meishan purebred lines alongside various crosses.

This dataset had been previously analysed in Lewis *et al.,* (2007, 2009a, 2009b). Herd breakdown for highly pathogenic PRRSV-2 (HP-PRRSV-2) (Tian *et al.,* 2007) had been observed at two time-points in August 2002 and October 2005, identified using regular ELISA testing. The precise start and end dates however were not clear.

These data consisted of 7,456 unique litter records detailing: sow identifier, sow line, sire of sow; sow of sow; sire used in fertilisation, services (the number of times the sow came into oestrus before successful insemination), matings (the number of inseminations before being successful), parity (the total number of litters the sow has had, including current litter, at the time of farrowing), net fostered (the net number of piglets fostered on or off the litter), as well as service date, farrowing date and weaning date, and the total numbers of piglets falling into one of four different reproductive outcomes.

These reproductive outcomes were recorded as counts of the number of piglets per litter born: *Mummified (Mum); Stillborn (Still); Alive;* and *Weaned* (*Wean*). Furthermore, three more trait values were derived: *Dead*, which was calculated as *Mum* + *Still* per litter; *Total observable foetuses (Tof)*, which was calculated as calculated as *Dead* + *Alive* per litter, and *Farrowing mortality (Fmor)* calculated by litter as *Dead*/*Tof* per litter. The *Dead* trait was created to capture the overall litter mortality without discriminating between the partly subjective classifications into mummified or stillborn foetuses. The term *Total observable*

foetuses (Tof) was used to reflect the fact that PRRSV can cause small foetuses which can be missed in the counts, as well as reabsorption of the foetus in utero. The trait *Fmor* considered the losses with respect to overall number of observed foetuses. In addition to this Gestation Length (*Gest*) was also calculated as Farrow Date minus Service Date in days. A summary table giving the definition of the traits is shown for reference in the appendix Table A.1.

The four, raw piglet counts by category (*Mum, Still, Dead, Alive*) and the four derived litter characteristics (*Tof*, *Fmor*, *Wean* and *Gest*) were used as the reproductive traits in the subsequent analyses.

The raw dataset comprised altogether 7,456 records between the periods from 13/05/2000 to 16/12/2006 by farrowing date. Parity had a range of 1-13, with between 1 and 1307 records for each parity class. There were relatively few (< 240) records with parity above eight. For this reason, parities of eight and higher were combined into parity class '8+'. For the 14 sow lines, there were between 1 and 2,042 litter records. Sow lines with less than 10 records were removed from the analyses (affecting two crosses and 10 records).14 further records were removed where no information was provided for number born alive, stillborn or mummified.

Figure 2.1 shows the number of farrow records per day throughout the 6-year period. Data prior to farrow date 03/06/2001 were sparsely populated with low numbers of records per day. These records were also excluded from the analyses to prevent bias in the 30-day cohort estimates used in the subsequent epidemic trend analysis. This led to the removal of a further 163 records. A plot showing a count of the number of litters recorded by farrow date in the retained records is shown in Figure 2.1.

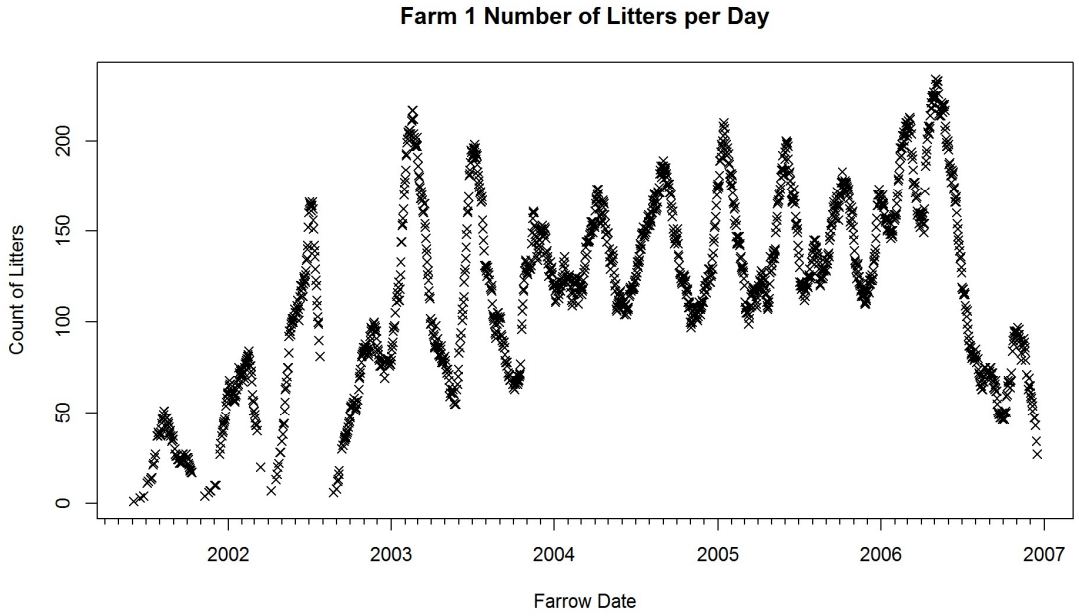

Figure 2.1 – Farm 1 Number of Litters Per Day
The number of litters recorded per day by farrow date.

This dataset used in the subsequent analyses comprised in total 7,031 records for 1,737 sows. Table 2.1 provides the frequency distribution of the number of records per sow in the cleansed dataset.

Table 2.1 – Frequency Distribution of the Number of Litters Recorded per Sow in the Cleansed Farm 1 Data

Number of Litters Class	Sow Frequency
1	387
2	336
3	196
4	136
5	156
6	149
7	108
8	91
9	110
10	56
11	2

Number of litter records per sow.

Farm 2

The second farm was a high-health Midwestern USA farm subject to monthly PRRSV antibody ELISA testing. This farm experienced one PRRSV-2 outbreak during which the farm was depopulated for decontamination and the animals slaughtered. Represented in farm 2 were 5 lines based on a cross of two lines in farm 1. One of these lines was from an identical breeding programme as for farm 1; the other crosses originated from a similar genetic background but with slightly different breeding objectives. The raw phenotypic data contained 11,053 records detailing: sow identifier, sow line, sire of sow; sow of sow; sire used in fertilisation, services (the number of times the sow came into oestrus before successful insemination), matings (the number of inseminations before being successful), as well as service date, farrowing date and weaning date, and the total numbers of piglets at farrowing categorised in the same manner as for farm 1. Identical to farm 1, for each litter the reproductive performance records providing numbers *Mummified* (*Mum*)*, Stillborn* (*Still*)*, Alive* and Weaned (*Wean*) were available, from which the derived traits (*Tof, Fmor Dead* and *Gest*), were calculated as for farm 1. Again forming the same reproductive traits in the subsequent analyses in this thesis.

For farm 2, the raw data did not contain direct information on the parity of sows, however did contain every successful farrowing of each sow. As such the parity of each sow at a given time point could be calculated as the number of prior successful farrows for each sow +1. The resulting data had a parity range of 1-13 matching the parity information on farm 1. Parities greater than 8 were combined to an aggregated grouping of eight plus, affecting 260 records. Also unavailable was the number of animals fostered on or fostered off the litter though it is known that some fostering took place. Given the impact that this has on the numbers weaned, and the potential for bias, the *Wean* trait was not assessed for Farm 2.

Similar to farm 1, the early time period in the dataset was sparsely populated and biased towards records of more mature sows with many parities. To prevent potential bias in the subsequent analyses, records prior to 01/11/2009 were discarded.

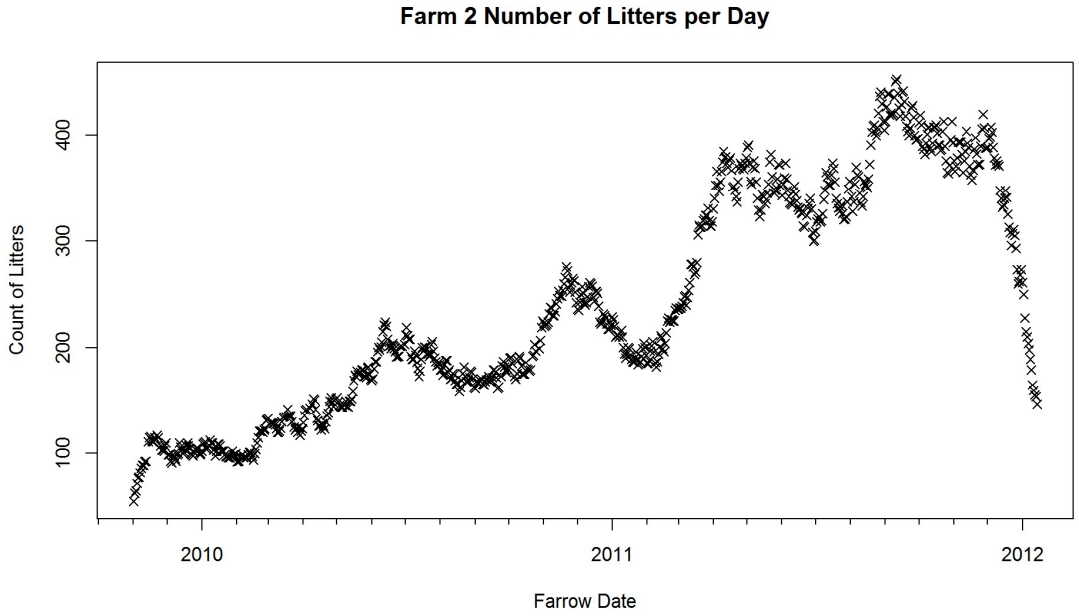

Figure 2.2 – Farm 2 Number of Litters Per Day
The number of litters recorded per day by farrow date.

The resulting dataset for farm 2 contained in total 6,487 records from 1,942 sows, the frequency distribution of which is shown in Table 2.2.

Table 2.2 – Frequency Distribution of the Number of Litters Recorded per Sow in the Cleansed Farm 2 Data

Number of Litters Class	Sow Frequency
1	295
2	553
3	199
4	227
5	427
6	241

Number of records per sow.

2.02.2 Data Partitioning into Epidemic and Non-Epidemic Phases

Data were partitioned into epidemic and non-epidemic phases according to a modification to the threshold-threshold method used in Lewis *et al.,* (2007). This method partitions the data by analysing the trends of the diverse reproductive performance traits using a 30-day rolling average. A mean for each trait under consideration is calculated at each farrow date, based on all records 15 days prior to, and 14 days subsequent to that farrow date (30 days total.). Hereafter, these 30-day trait means are referred to as the *trait 'trend*'. To partition these days into epidemic and non-epidemic phase, a baseline period is identified on each farm as a time-period where ELISA tests did not detect PRRSV. Within this period the 95th percentile of the trait trend is calculated as a threshold value. This threshold is then applied to the corresponding trait trend dataset to classify rolling average values according to whether they fall above/below this value. Confluent dates above this threshold coinciding with ELISA confirmed PRRSV outbreaks are then defined as *epidemic phase*.

Two weeks either side of each epidemic were extracted as a buffer to exclude data with low disease incidence rates. The corresponding data were not included in the analyses. The remaining data was defined as *non-epidemic phase* data.

This partitioning of the data is different in one aspect with regards to the partitioning used in (Lewis et al., 2009a). An emphasis was put on perinatal mortality described in terms of the raw counts and derived traits as defined. As such no lead/lag period was added to the epidemic period identified by the threshold to account for the gestation/lactation length of sows. This avoids the necessity to include periods exhibiting aberrant fertility outcomes associated with PRRS (such as number of services required for successful mating). Given perinatal mortality is expressed as an outcome following gestation whereas fertility-type traits are expressed prior to conception these two biologically different traits are observed at different times. Whilst in the previous study a lead time of one gestation length is added to disease phase to capture periods exhibiting aberrant fertility outcomes associated with PRRS,

focussing on perinatal mortality allows for a more stringent definition of disease phase pertinent to the specific traits under evaluation. It is noted that this precludes the analysis of some fertility outcomes included in Lewis et al., (2009a, 2009b) however during this period no evidence is seen of increases in prenatal and perinatal mortality (see results).

2.02.3 Statistical Analyses

Analyses were conducted on the two farms separately. All data processing and handling was done with R: A language and environment for statistical computing (R Foundation for Statistical Computing, (R Core Team, 2016). Mixed models were fitted using the lme4 and lmerTest package (Bates et al., 2016; Kuznetsova et al., 2016). Least square trait means associated with the fixed effects were calculated using the lsmeans package (Lenth, 2016).

The general mixed model, described in matrix notation is shown in Equation 2.1.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

Equation 2.1 – Matrix Equation of the Mixed Model

Where $\boldsymbol{y}$ is a vector of observations (the reproductive performance trait in consideration); $\boldsymbol{X}$ is a design matrix relating the observation to the fixed effects; $\boldsymbol{\beta}$ is a vector of fixed effects, $\mathbf{Z}$ is a design matrix relating the observations to the random effects, $\boldsymbol{u}$ is a vector of random effects, and $\boldsymbol{\varepsilon}$ is the error term. It was assumed that the random effects and errors were independently normally distributed with $u \sim N(0,G)$ and $\varepsilon \sim N(0,R)$ where G and R are the variance-covariance matrices with zero off diagonals

The significance of fixed effects was assessed using the F-test statistic and the significance of random effects was assessed using the Likelihood Ratio Test (LRT) (Neyman & Pearson, 1933; Self & Liang, 1987). Model fit was assessed based on the log-likelihood, the Akaike Information Criterion (AIC), and Root Mean Square Error (RMSE). The AIC provides an

estimate of model fit whilst penalising for the introduction of additional terms to generate accurate models whilst avoiding over-parameterisation (Akaike, 1974).

A stepwise elimination method was adopted to find the most comprehensive, but conservative model (in terms of over-fitting), pertinent to the trait under evaluation. This method initially includes all terms and possible interactions in a full model. Terms were removed from the model if they were not statistically significant based on the assessment methods described above ($p<0.05$), or did not decrease the AIC when dropped from the model. To ensure that models were hierarchically well formulated (HWF) lower order terms were retained when involved in higher order interactions (Kleinbaum *et al.,* 2008).

Fixed and random effects fitted in the models, and data transformation

The models were applied to the full datasets of both farms individually, comprising both epidemic and non-epidemic phases.

Models were generated for each of the above describe reproductive traits *(Mum, Still, Wean, Dead, Tof, Fmor, Gest*), not all of which were normally distributed. When model residuals were non-normally distributed, transformations were explored to ensure the models complied with the assumptions of the mixed model theory.

As repeated records were available for the majority of sows, sow was fitted as a random effect in all models. Sire of Sow, which was only available in farm 1 data, was also included in the models for farm 1 as random effect. Fixed effects included in the models comprised the binary trait 'Phase' to distinguish between epidemic and non-epidemic phase, as well as Parity, Sow Line and Year – Season, together with their interactions. In addition, *Tof* was considered as a fixed covariate in the models (where not considered as the response trait) to account for variation in the total number of piglets in the litter. Given fostering decisions were taken after the litter was farrowed, the number of net fostered individuals was also included as fixed covariate for the *Wean* trait in farm 1.

Models including just these basic components (and interactions) are termed the "Basic Model" hereafter.

Alterations to the Basic Model: Accounting for Variation in Exposure

In order to account for differences in the severity between different epidemics identified during partitioning, a unique identifier was allocated to each distinct epidemic period (epidemic ID) during the partitioning process and fitted as a fixed effect in addition to the basic model terms, thus replacing the binary trait 'Phase'. Non-epidemic phase was given an epidemic ID of zero. Models incorporating this effect are termed alternative model 1; or simply "Alt.1."

To additionally consider a dynamic within epidemic effect, a further model was established to try to account for potential changes occurring over the course of each epidemic. Given the trait trends generated in the partitioning process approximate the infected curve produced under the standard SIR model (see results), the rolling 30-day trend value specific to the date on which the litter farrowed was used as a covariate. This trait trend for the farrow date could be considered as an 'expected outcome' for the 30-day cohort to which the litter belonged, varying by date. In this model the trait trend value pertinent to the date of farrowing was fitted as a fixed covariate; to consider this an epidemic specific effect it was nested within epidemic ID.

The trend used corresponded to the trait under evaluation for the *Mum Still* and *Dead* traits. For the other traits (*Alive, Tof, Fmor, Wean* and *Gest*) all three loss trends were tested. The Dead trend consistently came out the most significant and as such was used in this model when analysing these traits. This model was termed alternative model 2 or simply, "Alt. 2."

2.03 Results

2.03.1 Data Partitioning

Whilst all loss trends were explored initially for partitioning, the slightly broader time-frame identified using the *Dead* trait trend (see appendix Figure A.1 to Figure A.6) captured the peaks or troughs for the trends in all reproductive traits. Therefore, the *Dead* trend was used to isolate epidemic periods for the subsequent analysis of all traits.

Farm 1

A plot of the *Dead* rolling average trend, identifying the initial baseline phase, 95th percentile threshold value and highlighting periods above and below the threshold, is shown Figure 2.3.

For farm 1 three prolonged periods above the threshold were identified coinciding with ELISA positive results for PRRSV, as shown by the three red periods with increased number of dead piglets in Figure 2.3. The exact corresponding periods were from 25/08/2005 to 16/12/2002, labelled Epidemic 1; from 09/11/2005 to 19/03/2006, labelled Epidemic 2; and from 17/07/2006 to 16/09/2006, labelled Epidemic 3 (Figure 2.3).

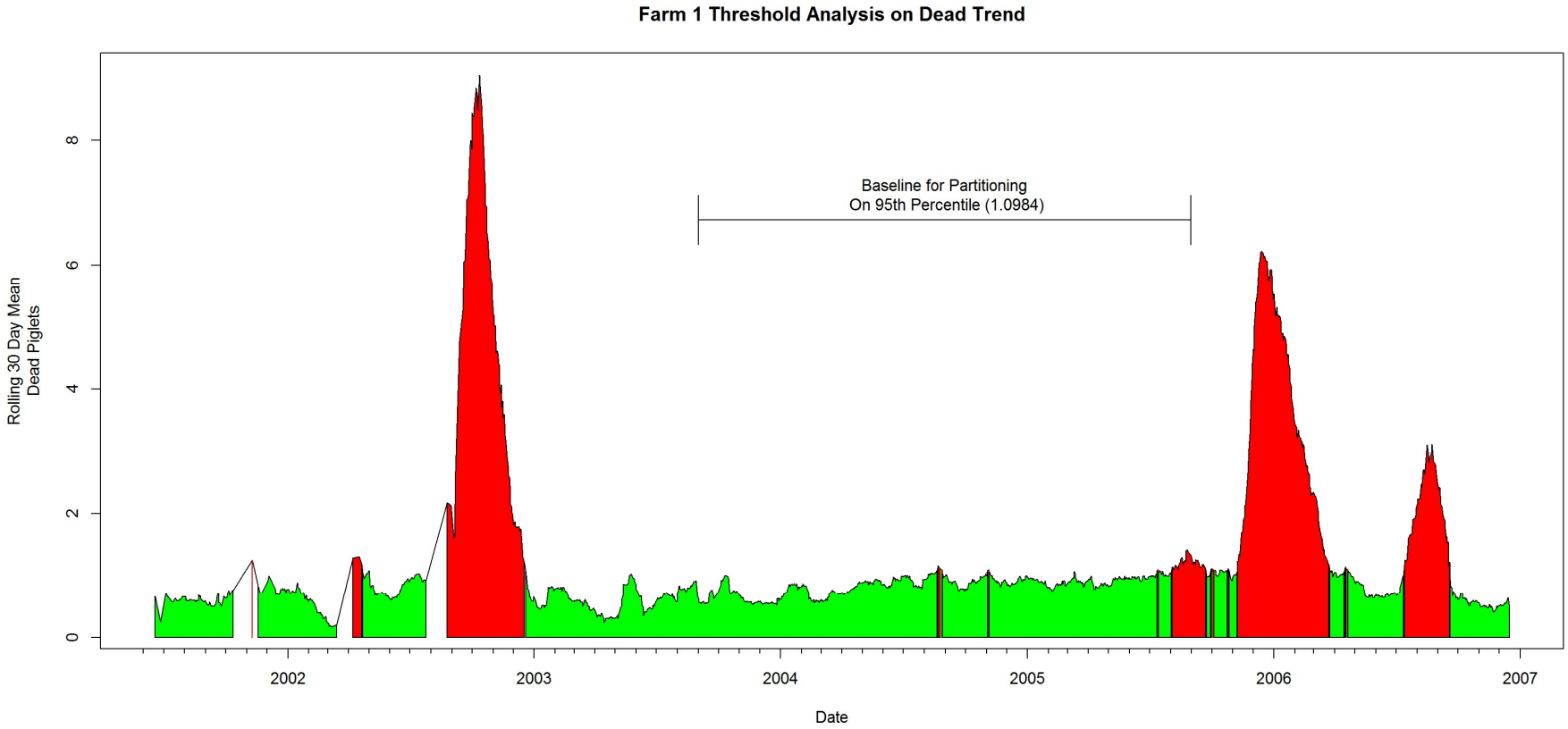

Figure 2.3 – Farm 1 Dead Trend Threshold Analysis

Periods corresponding to values within the 95th percentile of the calculated trend within the baseline period (bracketed) are shown in green, and periods corresponding to the remaining 5% values are shown in red.

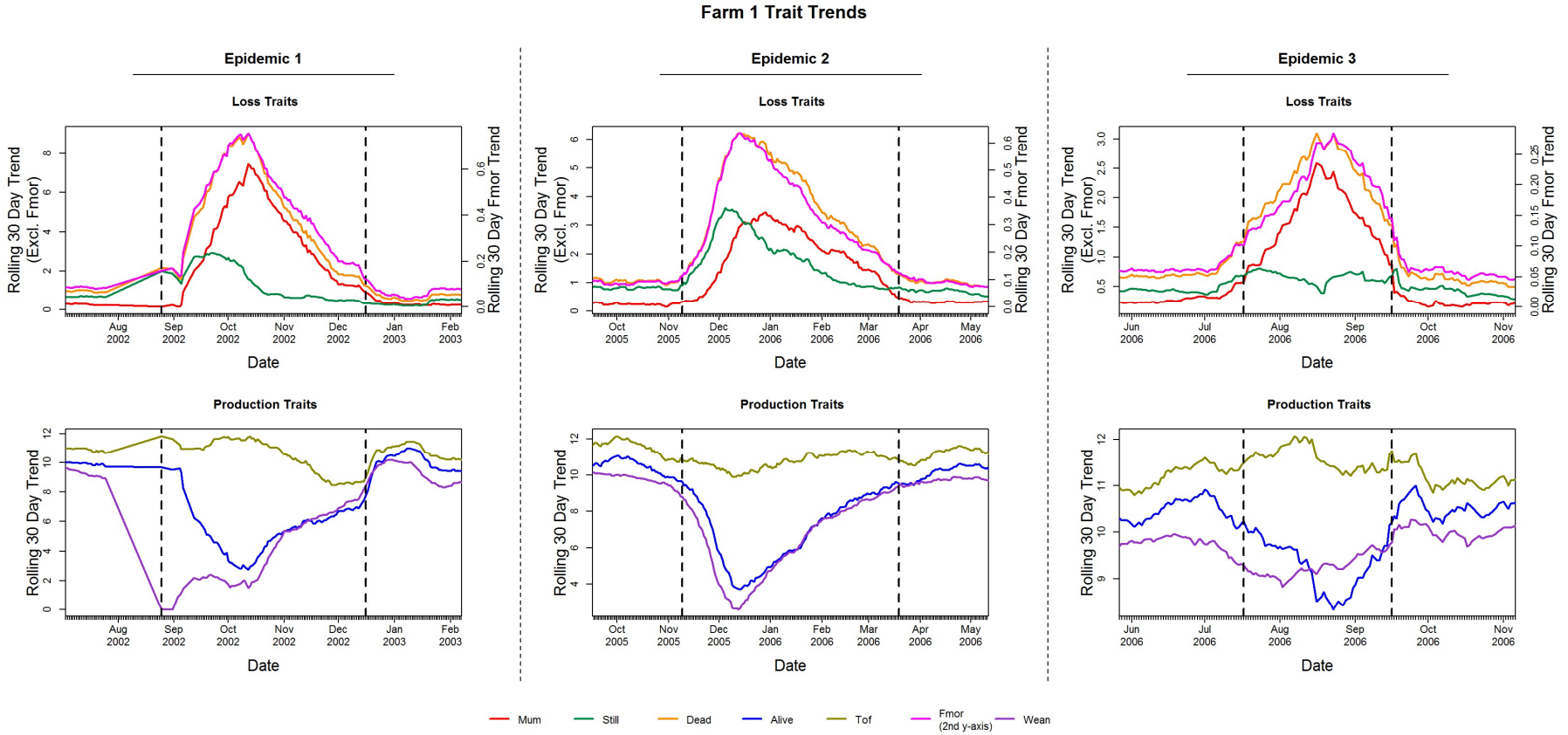

Figure 2.4 – Trait Trends for Farm 1 Epidemics

Loss traits (*Mum, Still, Dead* and *Fmor*) shown top, production traits (*Alive Tof* and *Gest*) shown bottom due to differences of scale. Partitioning dates, calculated using the *Dead* trend threshold are shown as dashed vertical lines.

Figure 2.4 shows trait trends for the epidemics identified on Farm 1 using the *Dead* trait trend threshold partitioning, vertical dotted lines marking the periods identified. Note an apparent aberration in the trends for epidemic 1 in August 2002, this coincides with no recorded litters farrowed in this period (see Figure 2.1).

Table 2.3 shows the number of sows and records per sow for the 4 different pure-bred sow lines, and the various cross-breeds, for the non-epidemic and epidemic phases. All lines were represented in both phases, but the non-epidemic phase comprised 4.99 times more records than the epidemic phases together.

Table 2.3 – Number of Animals By Sow Line in the Epidemic and Non-Epidemic Phases.

Breed	Non-Epidemic		Epidemic	
	Number of Sows	Number of Records	Number of Sows	Number of Records
Landrace	397	1310	302	320
Large White	433	1546	284	311
Large White From NPD	364	1494	213	225
Synthetic (Duroc × White)	133	392	86	93
Pietrain (Hal+)	29	83	16	16
Meishan	61	274	34	37
Synthetic (Large White × White)	23	120	7	7
Synthetic (Large White x Synthetic Sire Line)	6	18	9	9
Pietrain (Hal+ × Hal-)	12	16	8	9
Synthetic (White × Meishan)	15	25	20	22
Unknown	4	14	3	4
Synthetic (White × Duroc)	15	15	9	9

The exact dates identified under the trend analysis, baseline and threshold values for other traits are shown in the appendix Figure A.1 to Figure A.3

Farm 2

For Farm 2, two extended periods above the baseline threshold were observed (Figure 2.5), one from the end of December 2010 to the end of March 2010, the other from mid-December 2011 to mid-January 2012. Only the period from mid-December 2011 to mid-January 2012 coincided with an ELISA positive PRRSV test result. The exact corresponding periods identified under the trend analysis were 11/12/2001 to 14/01/2012 labelled epidemic 4 shown in Figure 2.6.

The other period corresponding to dates 29/12/2009 to 30/03/2010 was labelled Epidemic 5, treated as an unconfirmed epidemic, and not included in the analysis, unless explicitly stated. The reproductive trait trends in both epidemic periods (confirmed and unconfirmed) are shown in Figure 2.5.

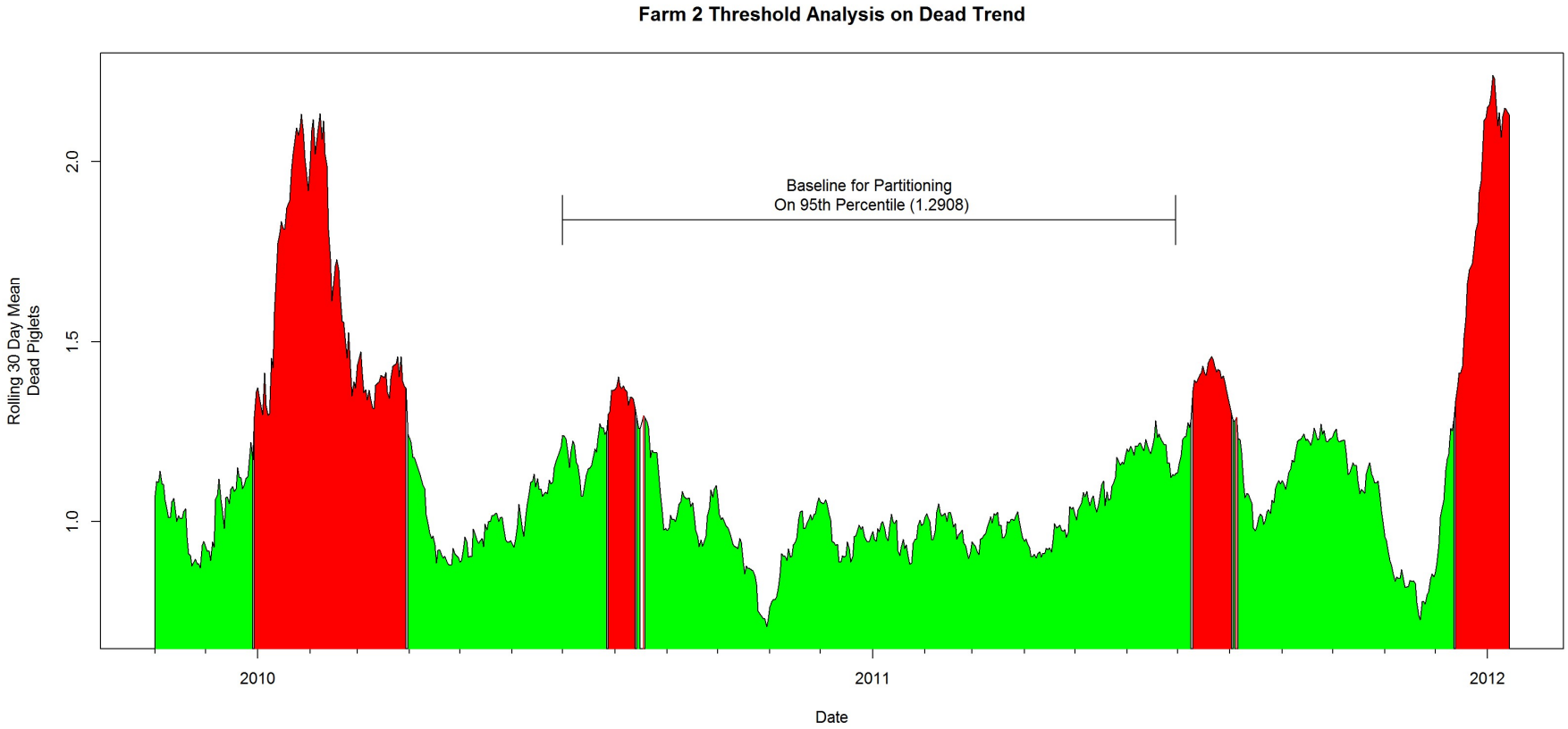

Figure 2.5 – Farm 2 Dead Trend Threshold Analysis

Periods corresponding to values within the 95th percentile of the calculated trend within the baseline period (bracketed) are shown in green, and periods corresponding to the remaining 5% values are shown in red.

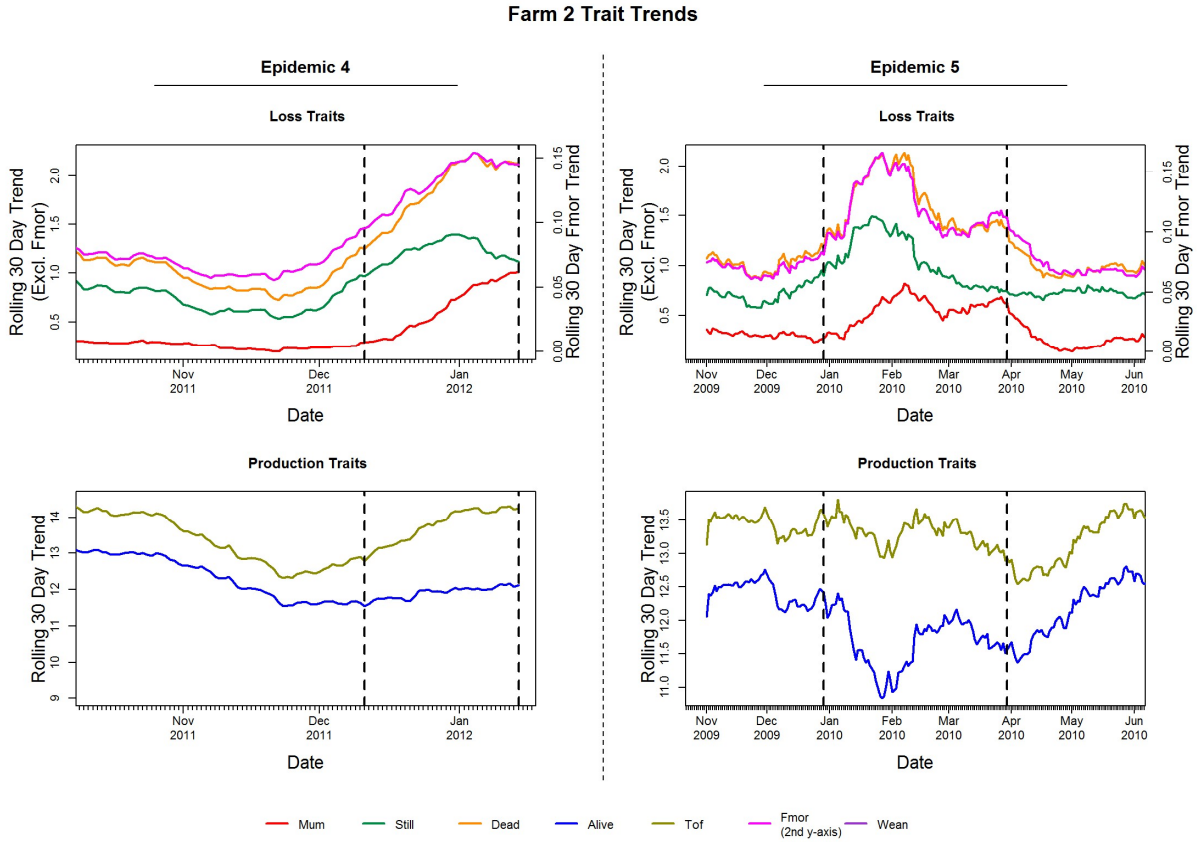

Figure 2.6 – Trait Trends for Farm 2 Epidemic 4

Loss traits (*Mum, Still, Dead* and *Fmor*) shown left, production traits (*Alive Tof* and *Gest*) shown right for clarity due to differences of scale. Partitioning dates, calculated using the *Dead* trend threshold are shown as dashed vertical lines.

The exact dates identified under the trend analysis, baseline and threshold values for other traits are shown in the appendix Figure A.4 to Figure A.6

Qualitative description of the impact of PRRSV outbreaks on reproductive performance traits

The trend analysis showed a clear increase in loss trait (*Mum, Still, Dead*) and *Fmor* trends and a clear decrease in two production traits (*Alive* and *Wean*) trends over the epidemic periods (Figure 2.4) No clear, systematic, pattern was observed in *Tof* trends. There are some exceptions to this. Epidemic 3 on farm 1 did not experience a noticeable increase in the *Still* trend and Epidemic 5 on farm 2 did not experience a noticeable decrease for the *Alive* trend. Interestingly the trend analysis identified a pattern in the traits whereby the increase or decrease in each trait trend did not occur concurrently but appeared offset with regards to each other.

Trait trends for Epidemic 1 are shown in Figure 2.4. At the early stage of the epidemic the trend appears more erratic than for Epidemic 2 for all traits except *Alive*. Note that the first effect of the epidemic is reflected in the *Wean* trend which crashes abruptly to 0. At the same time as the *Wean* trend approaches zero there is a large gap with no recorded farrows between dates 25/07/2002 and 25/08/2002 (shown in Figure 2.1). The epidemic would conceivably first show in reproductive data in the number weaned as late gestation\recently farrowed piglets succumbed to the virus, followed by an absence of any reported farrowing where sows infected earlier in gestation aborted the litter. A gradual recovery in the *Wean* trend and in number of farrows is observed throughout the epidemic. In Figure 2.4 the losses are observed in a rise in the *Still* trend followed by the *Mum* trend. This pattern is clearly reflected in all the PRRSV confirmed epidemic trends on both farms, except for farm 1, epidemic 2 where the *Still* trend did not show a noticeable increase. A similar pattern is observed in the unconfirmed epidemic on farm 2 which shows patterns in all the trends sequentially.

2.03.2 Statistical Analysis

Farm 1

Trait means including Standard Error of the Mean (SEM), sow and record counts are shown for the non-epidemic and epidemic phases (including by epidemic breakdown) for farm 1 data in Table 2.4.

Table 2.4 – Summary statistics showing Counts and Trait Means for Farm 1

Level	Records	Sows	*Mum*	*Still*	*Dead*	*Alive*	*Tof*	*Fmor*	*Wean*	*Gest*
All	6,369	1,708	0.55 (0.02)	0.73 (0.02)	1.28 (0.03)	9.76 (0.04)	11.04 (0.04)	0.12 (0.002)	9.13 (0.03)	115.59 (0.02)
Non-epidemic Phase	5,307	1,492	0.23 (0.01)	0.592 (0.01)	0.82 (0.02)	10.27 (0.04)	11.09 (0.04)	0.07 (0.002)	9.6 (0.03)	115.6 (0.02)
Epidemic Phase	1,062	991	2.18 (0.10)	1.42 (0.07)	3.60 (0.12)	7.19 (0.13)	10.79 (0.10)	0.33 (0.01)	6.77 (0.10)	115.52 (0.07)
Epidemic 1	199	199	3.46 (0.31)	1.03 (0.12)	4.49 (0.34)	5.51 (0.28)	10.01 (0.25)	0.40 (0.03)	5.03 (0.24)	115.23 (0.18)
Epidemic 2	705	705	1.95 (0.11)	1.68 (0.09)	3.64 (0.14)	7.18 (0.16)	10.82 (0.12)	0.34 (0.01)	6.70 (0.12)	115.73 (0.0856)
Epidemic 3	158	158	1.57 (0.24)	0.72 (0.11)	2.29 (0.27)	9.35 (0.34)	11.64 (0.24)	0.2 (0.02)	9.23 (0.12)	114.99 (0.17)

Summary statistics for data under different partitions showing number of records, number of sows and means with standard error of the mean for the reproductive performance traits considered in this study.

Consistent with previous findings by Lewis *et al.,* (2009a, 2009b), the Log_{10}(Trait+1) transformation was found to be required for the loss count traits *Mum, Still* and *Dead*. Transformations were not required, however, for the (more normally distributed) *Alive, Tof, Fmor, Wean* and *Gest* traits.

The final basic models (not accounting for differences in exposure during the epidemic phase) for the diverse reproductive traits differed, depending on the significance of the fixed effects. Both random effects sow and sire of sow were found to be statistically significant according to the LRT.

Table 2.4 shows the p-values for all significant, retained, fixed effects and interactions. Year season was found not to be significant for any of the traits, as found by Lewis *et al.,* (2009a) and was dropped. Parity, in keeping with previously reported results was highly significant across traits and datasets. *Tof* too was highly significant where fitted except for trait *Fmor*. When *Tof* was included as a fixed covariate in the models, sow line was non-significant for the loss type traits *Mum, Still, Dead* and *Fmor*. Sow line was however significant for these traits if *Tof* was not included as a covariate.

In order to explore the impact of differences in the severity of each epidemic on trait means, epidemic ID was incorporated in the subsequent set of models for each trait, in place of the 'phase' in the basic model (Alternative Model 1 or Alt.1).

Table 2.5 – Farm 1 F-Test Statistical Significance for All Fixed Effects, Covariates and Interactions Under the Basic and Alternative Models For All Traits

Model	Trait	Parity	*Tof*	Sow Line	Epidemic†	Epidemic† × Trend	Epidemic† × Parity	Epidemic† × *Tof*	Epidemic† × Sow Line × Cycle	Net Fostered	Epidemic† × Net Fostered	Epidemic† × *Tof* × Net Fostered
Basic	*Mum*¥	***	***	—	***	—	***	***	—	—	—	—
	Still¥	***	***	—	*	—	***	*	—	—	—	—
	Dead¥	***	***	—	***	—	***	***	—	—	—	—
	Alive	***	***	***	***	—	—	—	***	—	—	—
	Tof	***	—	***	*	—	—	—	**	—	—	—
	Fmor	***	—	—	***	—	***	—	—	—	—	—
	Wean	***	***	—	***	—	***	—	—	***	***	***
	Gest	—	***	—	***	—	—	—	—	—	—	—
Alt.1	*Mum*¥	***	***	—	***	—	***	***	—	—	—	—
	Still¥	***	***	—	**	—	***	***	—	—	—	—
	Dead¥	**	***	—	***	—	***	***	—	—	—	—
	Alive	***	***	***	***	—	—	—	***	—	—	—
	Tof	***	—	***	***	—	—	—	***	—	—	—
	Fmor	***	—	—	***	—	***	—	—	—	—	—
	Wean	***	***	—	***	—	***	—	—	***	***	***
	Gest	—	***	—	***	—	—	—	—	—	—	—
Alt.2	*Mum*¥	***	***	—	***	***	***	***	—	—	—	—
	Still¥	***	***	—	***	***	***	***	—	—	—	—
	Dead¥	*	***	—	**	***	***	***	—	—	—	—
	Alive	*	***	***	*	***	—	—	***	—	—	—
	Tof	***	—	***	***	***	—	—	***	—	—	—
	Fmor	***	—	—	***	***	***	—	—	—	—	—
	Wean	*	***	—	***	***	***	—	—	***	***	***
	Gest	*	***	—	***	***	—	—	—	—	—	—

¥Trait transformed by log(trait+1), †When fitted in the basic model this effect is the Phase term, when fitted in the alternative models this is the Epidemic ID term, ‡The fitted trend is the 30 day rolling mean corresponding to the trait for *Mum, Still, Dead*. for all other traits the *Dead* Trend was used. Significance: *** P<0.001; ** P< 0.01; * P<0.05; — Not significant and/or not included in model (see text).

The epidemic effect, regardless of how it was included in the models, was highly significant across traits and models. Significant too is the trend by epidemic ID interaction, indicating that differences in exposure over time, as represented by the rolling average trend of the trait in consideration (or *Dead* trait) explained a significant proportion of the observed variation in reproductive performance.

Looking at model fit, Table 2.4 shows some measures of the relative quality of the models across models and traits, including the AIC, root mean square error, the log- likelihood as well as the significance of the sow and sire of sow effects.

Table 2.6 – Farm 1 Estimates of Model Fit, and Random Effect LRT p-value (P)

Trait	Model	Akaike Information Criterion	Log Likelihood	Root Mean Square Error	Sire LRT P	Sow LRT P
Mum¥	Basic	7863.50	-3910.75	0.433	0.165	<0.001
	Alt.1	7552.08	-3742.04	0.423	0.21	0.008
	Alt.2	7019.44	-3471.72	0.405	0.193	0.006
Still¥	Basic	8955.50	-4456.75	0.463	<0.001	<0.001
	Alt.1	8952.84	-4442.42	0.461	<0.001	<0.001
	Alt.2	8529.84	-4226.92	0.447	<0.001	<0.001
Dead¥	Basic	11021.12	-5489.56	0.546	<0.001	<0.001
	Alt.1	10833.45	-5382.73	0.538	0.003	<0.001
	Alt.2	10359.19	-5141.60	0.519	0.004	<0.001
Alive	Basic	26114.04	-12916.02	1.744	<0.001	0.003
	Alt.1	25732.45	-12680.23	1.703	<0.001	0.40
	Alt.2	25052.23	-12336.12	1.627	<0.001	0.36
Tof	Basic	31097.42	-15408.71	2.432	<0.001	<0.001
	Alt.1	30949.22	-15289.61	2.416	<0.001	<0.001
	Alt.2	30938.13	-15280.06	2.41	<0.001	<0.001
Fmor	Basic	-3871.44	1954.72	0.165	<0.001	<0.001
	Alt.1	-4119.01	2089.50	0.163	<0.001	<0.001
	Alt.2	-4941.647	2504.823	0.154	<0.001	0.002
Wean	Basic	26851.083	-13401.54	1.873	0.002	<0.001
	Alt.1	26479.913	-13200.96	1.852	0.009	0.03
	Alt.2	25709.763	-12811.88	1.75	0.022	0.05
Gest	Basic	23543.865	-11630.93	1.18	<0.001	<0.001
	Alt.1	23402.977	-11515.489	1.158	<0.001	<0.001
	Alt.2	23305.214	-11462.607	1.146	<0.001	<0.001

¥Transformed log(trait + 1).

With few exceptions, models including random sow and sire of sow effects provided in general a significantly better model fit than models without these effects. Sire of sow did not seem to significantly affect only the number of mummified piglets (Table 2.6).

Consistently from the basic model to the alternative model 1 and alternative Model 2 the model fit improved, shown by reductions in both the AIC and RMSE, and by significant improvement in the log-likelihood (p-value of LRT was <0.001 for all traits).

Figure 2.7 shows the least squares means for the *Mum* trait for all models. Estimates are provided for the epidemic effect (phase or epidemic ID depending on what is fitted in the model), with standard error and 95% confidence level. These fix covariates at their mean level values and average over the other terms in the model. For all three types of models, a clear distinction is seen between the non-epidemic phase and the epidemic phase \ epidemics. LSM *Mum* values are significantly lower during the non-epidemic phase ($p<0.001$). Furthermore, Figure 2.7 shows that there is a significant difference in the LSM *Mum* values between the individual epidemics, although all individual epidemics resulted in significantly higher *Mum* LSMs than observed during the non-epidemic periods. The LSM estimates and confidence intervals were very similar between the alternative models 1 and 2.

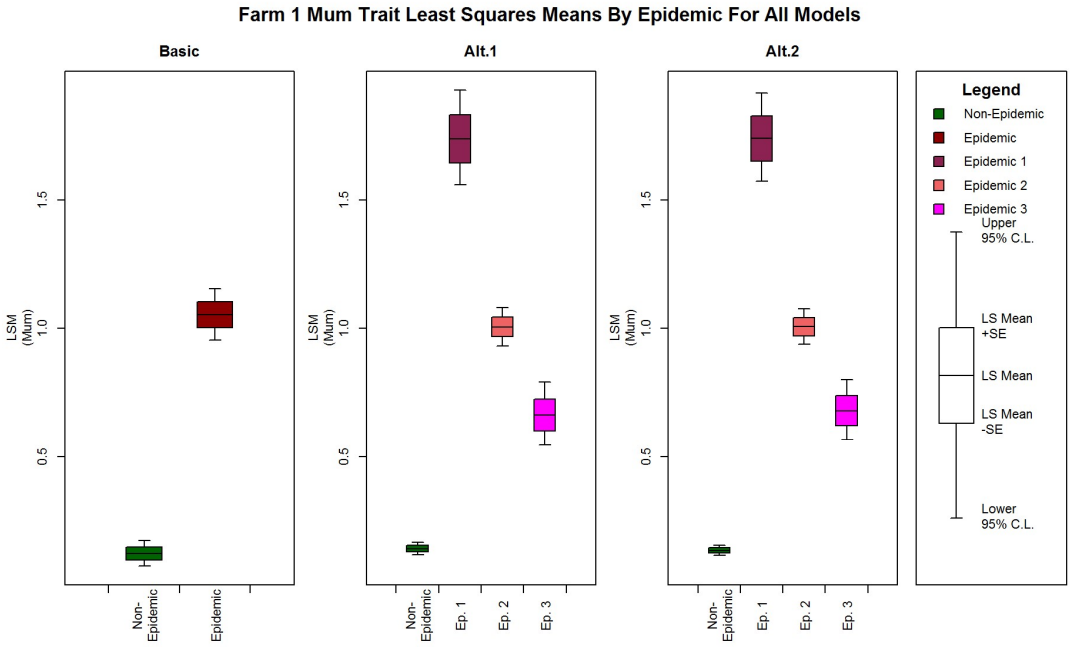

Figure 2.7 – Farm 1 Mum Trait Least Squares Means for Epidemic Effect Estimated Using Different Models

Log(*Mum*+1) trait fitted, back-transformed onto the response scale. Covariates involved in interactions with the epidemic factor are estimated at their mean for each epidemic level.

Figure 2.8 shows the LSM estimates for all reproductive performance traits estimated with the basic model. Statistically significant differences in the LSMs associated with the epidemic and non-epidemic phase are seen for all traits, except for *Tof* and *Gest,* indicating that *Mum*, *Still*, *Dead*, *Alive Fmor* and *Wean* present useful, quantitative indicators of disease impact for further analysis.

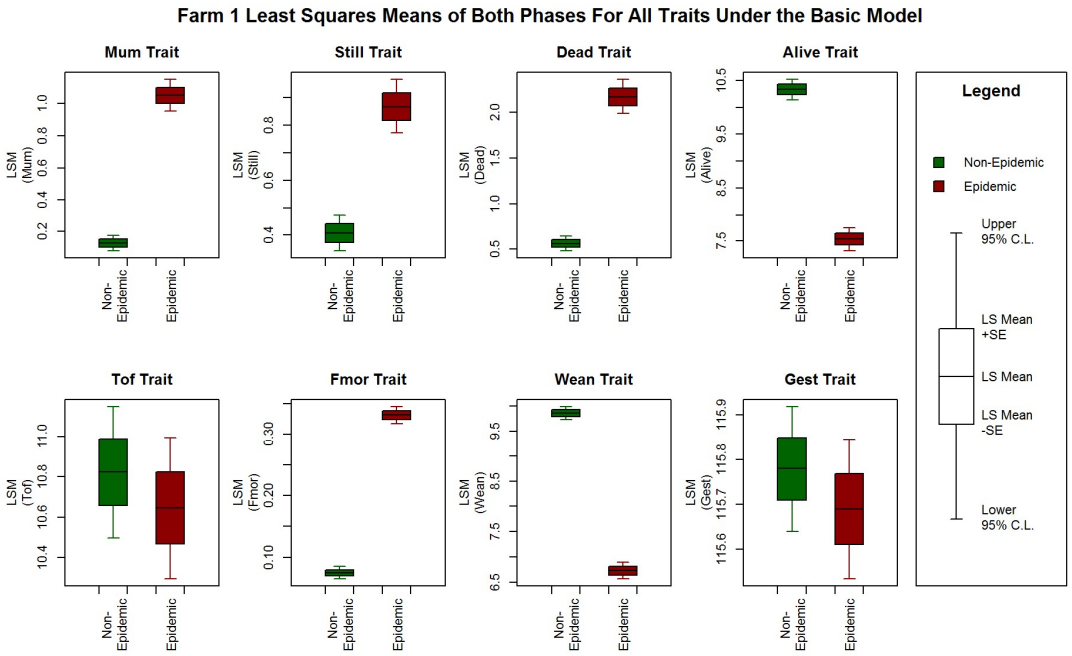

Figure 2.8 – Farm 1 Least Squares Means for All Traits by Phase Estimated Under the Basic Model

Mum, Still and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

In addition to the epidemic characteristics, the sow parity by epidemic interaction was also found significant in the statistical models for the majority of reproductive traits. The parity effect on the diverse reproductive traits for the epidemic and non-epidemic phases estimated with the basic model are shown in Figure 2.9. Except for *Tof* and *Gest*, all traits show a significant loss in reproductive performance associated with the PRRSV outbreak across all parities. However, whereas reproductive performance of sows reduces systematically in later parities in the absence of a PRRSV outbreak, this systematic effect seems to disappear

during the epidemic period. This could be partly an artefact of the limited amount of available records per parity during the epidemic period.

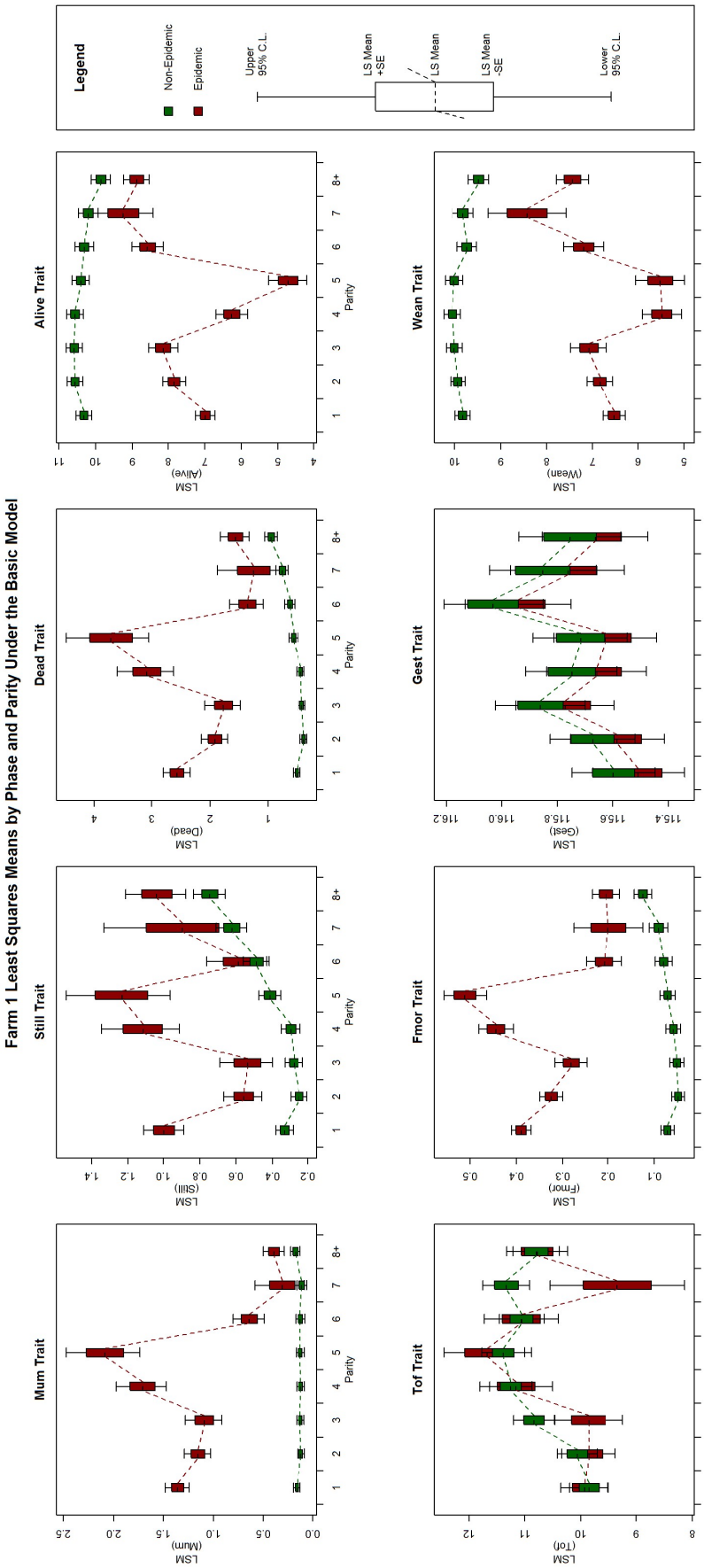

Figure 2.9 – Least Squares Means by Phase and Parity for All Traits Estimated Under the Basic Model for Farm 1

Mum, *Still* and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

Figure 2.10 shows the non-systematic parity effect on reproductive performance during the epidemic period observed for the basic model, however at the individual epidemic level this may not be the case. In epidemic 1, the most severe epidemic, the number of mummified piglets increased significantly at later parities. For epidemics 3 this pattern is reversed, whereas in epidemic 2 the non-systematic effect it retained. This was true for both models, Alt.1 and Alt.2 (Figure 2.10).

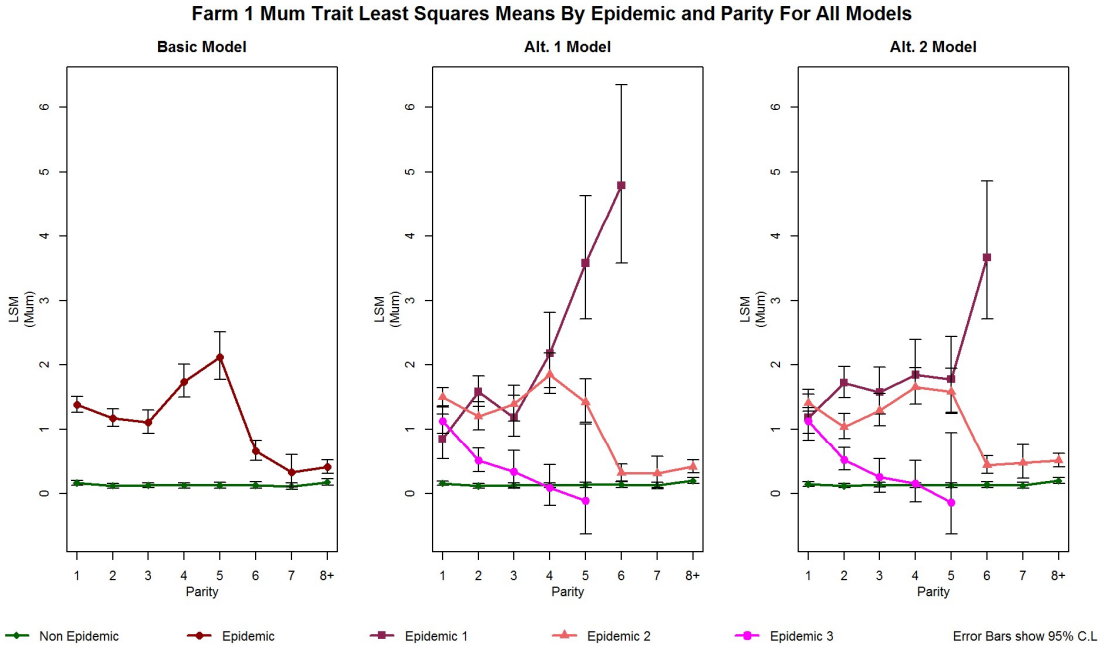

Figure 2.10 – Farm 1 Mum Trait Least Squares Means by Parities and Epidemic Factor Level Included in Model

Log(*Mum*+1) trait fitted, back-transformed onto the response scale. Covariates involved in interactions with the epidemic factor are estimated at their mean for each epidemic level.

Farm 2

Trait means including SEM, sow and record counts are shown for the non-epidemic and epidemic phase for farm 2 data in Table 2.7.

Table 2.7 – Summary statistics showing Counts and Trait Means of Traits and Covariates for Farm 2 Cleansed Dataset

Level	Records	Sows	*Mum*	*Still*	*Dead*	*Alive*	*Tof*	*Fmor*	*Wean*	*Gest*
All	6,069	1,914	0.31 (0.01)	0.84 (0.02)	1.14 (0.02)	12.57 (0.04)	13.71 (0.04)	0.08 (0.001)	10.54 (0.03)	115.62 (0.02)
Non-epidemic Phase	5,386	1,877	0.27 (0.01)	0.79 (0.02)	1.07 (0.02)	12.66 (0.04)	13.72 (0.04)	0.07 (0.001)	10.79 (0.02)	115.63 (0.02)
Epidemic Total	683	627	0.59 (0.05)	1.16 (0.06)	1.75 (0.08)	11.88 (0.13)	13.63 (0.12)	0.13 (0.006)	8.62 (0.14)	115.47 (0.06)
Epidemic 4	339	339	0.62 (0.09)	1.28 (0.09)	1.90 (0.13)	12.06 (0.19)	13.96 (0.18)	0.13 (0.009)	7.84 (0.24)	115.77 (0.09)
Epidemic 5	344	344	0.57 (0.07)	1.04 (0.08)	1.61 (0.11)	11.70 (0.18)	13.31 (0.17)	0.12 (0.008)	9.39 (0.15)	115.18 (0.08)

Summary statistics for data under different partitions showing number of records, number of sows and means with standard error of the mean for the reproductive performance traits considered in this study.

The investigation into reproductive performance was initially confined to the data excluding the unknown epidemic. All models included random effect of sow but due to the lack of pedigree information no sire effect could be fitted. Consistent with previous findings, in Lewis *et al.,* (2009a, 2009b) Log_{10}(Trait+1) transformation was found to be required for the loss count traits of *Mum*, *Still* and *Dead* and not required, however, for the *Alive*, *Tof, Fmor* and *Wean* and *Gest* Traits.

Basic models were generated in the same manner as for farm 1. Given the lack of multiple epidemics alternative model 1 has no additional contrasts as compared to the basic model (both the Phase and Epidemic Id are two level factors). The equivalent for model 2 was generated with an epidemic ID × trend term to the basic model terms.

The final basic models (not accounting for differences in exposure during the epidemic phase) for the diverse reproductive traits differed, depending on the significance of the fixed effects. Both random effects sow and sire of sow were found to be statistically significant according to the LRT.

Table 2.8 shows the p-values for all significant, retained, fixed effects and interactions.

Table 2.8 – Farm 1 F-Test Statistic P-Values for All Fixed Effects, Covariates and Interactions Under the Basic and Alternative 2 Models For All Traits

Model	Trait	Parity.	Sow Line.	Tof	Epidemic†	Epidemic† × Trend‡	Epidemic × Parity.	Epidemic† × Sow Line.
Basic	*Mum*¥	**	•	***	***	—	***	—
	Still¥	***	***	***	***	—	—	—
	Dead¥	***	***	***	***	—	*	—
	Alive	***	•	***	※	—	—	*
	Tof	***	***	—	※	—	—	•
	Fmor	***	***	—	***	—	***	—
	Gest	***	—	***	—	—	—	—
Alt.2	*Mum*¥	•	•	***	※	***	***	—
	Still¥	***	***	***	※	***	—	—
	Dead¥	***	***	***	※	***	*	—
	Alive	***	※	***	※	***	—	•
	Tof	***	***	—	※	***	—	•
	Fmor	***	***	—	※	***	***	—
	Gest	***	—	***	※	***	—	—

¥Trait transformed by log(trait+1), †When fitted in the basic model this effect is the Phase term, when fitted in the alternative model this is the Epidemic Id term, ‡The fitted trend is the 30 day rolling mean corresponding to the trait for *Mum, Still, Dead* for all other traits the *Dead* Trend is used, Significance: *** P<0.001; ** P< 0.01; * P<0.05; • P<0.1; ※ Not significant but retained in model; — Not significant and not included in model.

A significant epidemic effect (phase or epidemic ID × trend interaction) is seen in all traits except *Gest*. For the *Alive* and *Tof* traits this effect is only shown as significant in the epidemic × sow line interaction. Fewer characteristics were found to have significant epidemic interaction than on farm 1, including the epidemic × parity interaction. This could either be as a result of a more homogenous population in farm 2 as compared to farm 1 in terms of breed varieties (providing less contrast in these factors), or simply a product of the reduced number of records.

Parity and *Tof* are again significant across all the traits tested. Sow line is also significant for all the traits except *Gest*. Despite (as before) accounting for the litter size by fitting the covariate *Tof*.

Table 2.9 shows some measures of the relative quality of the models across models and traits, including the AIC, root mean square error, the log-likelihood as well as the significance of the sow effect.

Table 2.9 – Farm 2 Estimates of Model Fit, and Random Effect LRT p-value (P)

Trait	Model	Akaike Information Criterion	Log Likelihood	Root Mean Square Error	Sow LRT P
Mum¥	Basic	4240.43	-2097.21	0.33	<0.001
	Alt. 2	4170.99	-2060.49	0.33	<0.001
Still¥	Basic	7849.03	-3908.52	0.44	<0.001
	Alt. 2	7803.38	-3883.69	0.44	<0.001
Dead¥	Basic	8723.96	-4338.98	0.48	<0.001
	Alt. 2	8672.57	-4311.29	0.47	<0.001
Alive	Basic	19411.43	-9685.72	1.22	<0.001
	Alt. 2	19346.44	-9651.22	1.21	<0.001
Tof	Basic	28813.70	-14387.85	2.51	<0.001
	Alt. 2	28796.86	-14377.43	2.51	<0.001
Fmor	Basic	-10501.11	5272.56	0.09	<0.001
	Alt. 2	-10559.57	5303.79	0.09	<0.001
Gest	Basic	18057.88	-9016.94	0.82	<0.001
	Alt. 2	18009.63	-8990.82	0.81	<0.001

¥Transformed log(trait + 1)

Models including the sow effect show considerably improved model fit than models without these effects. Under all measures of model fit alternative model 2 performs better than the basic model including significant improvements to the LRT between models p value $<1.49\times10^{-5}$ for all traits (Table 2.9).

Figure 2.11 shows the least squares means for the *Mum* trait for both models. Estimates are provided for phase, with standard error and 95% confidence level. For both models a clear distinction is seen between the two phases. LSM *Mum* values are significantly lower during the non-epidemic phase ($p<0.001$). The LSM estimates and confidence intervals are very similar between the basic model and alternative model 2.

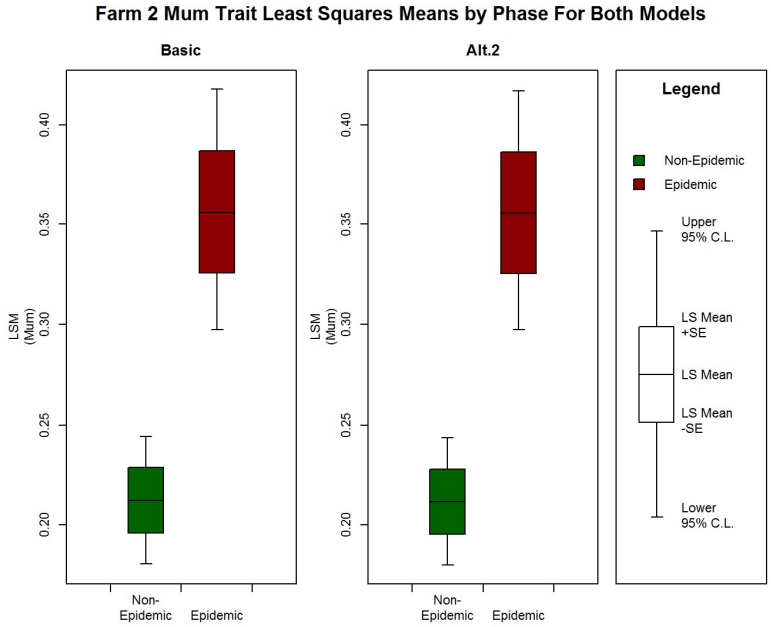

Figure 2.11 – Farm 2 Mum Trait Least Squares Means for Epidemic Effect Estimated Using Different Models

Log(*Mum*+1) trait fitted; back-transformed onto the response scale. Covariates involved in interactions with the epidemic factor are estimated at their mean for each epidemic level.

Figure 2.12 shows the LSM estimates for all reproductive performance traits estimated with the basic model. Statistically significant differences in the LSMs associated with the epidemic and non-epidemic phase are seen for all traits, except for *Tof* and *Gest,* indicating that *Mum*, *Still*, *Dead*, *Alive* and *Fmor* present useful, quantitative indicators of disease impact for further analysis.

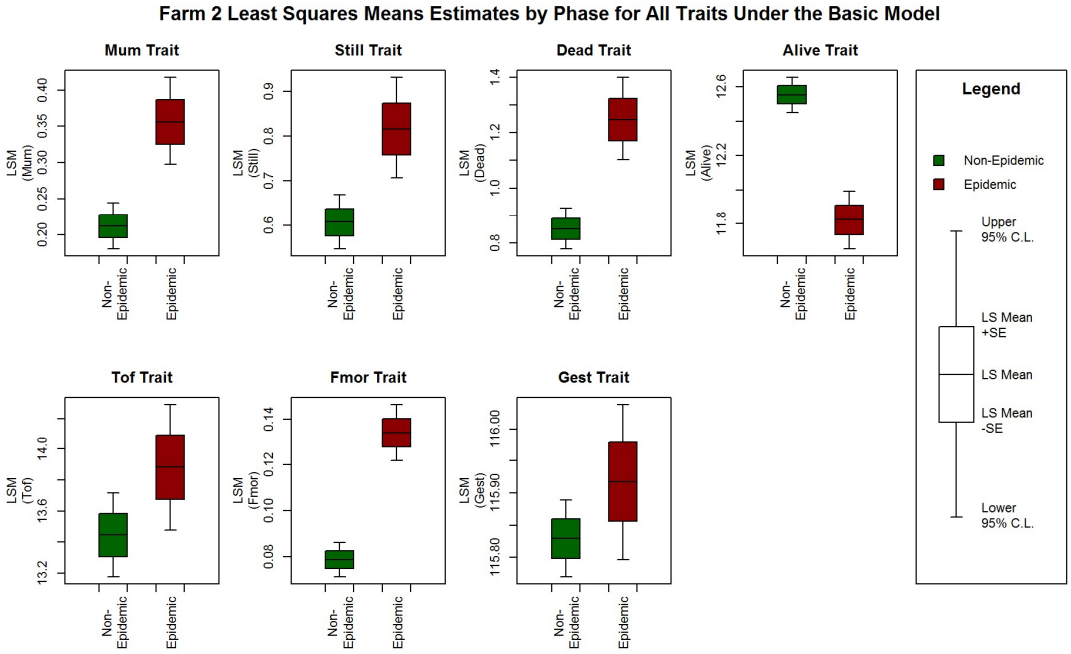

Figure 2.12 – Farm 2 Least Squares Means for All Traits by Phase Estimated Under the Basic Model

Mum, *Still* and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

Fewer traits showed a significant parity by epidemic effect on farm 2, these traits are *Mum, Dead,* and *Fmor*. The LSM estimated under the basic model for these traits is shown Figure 2.13.

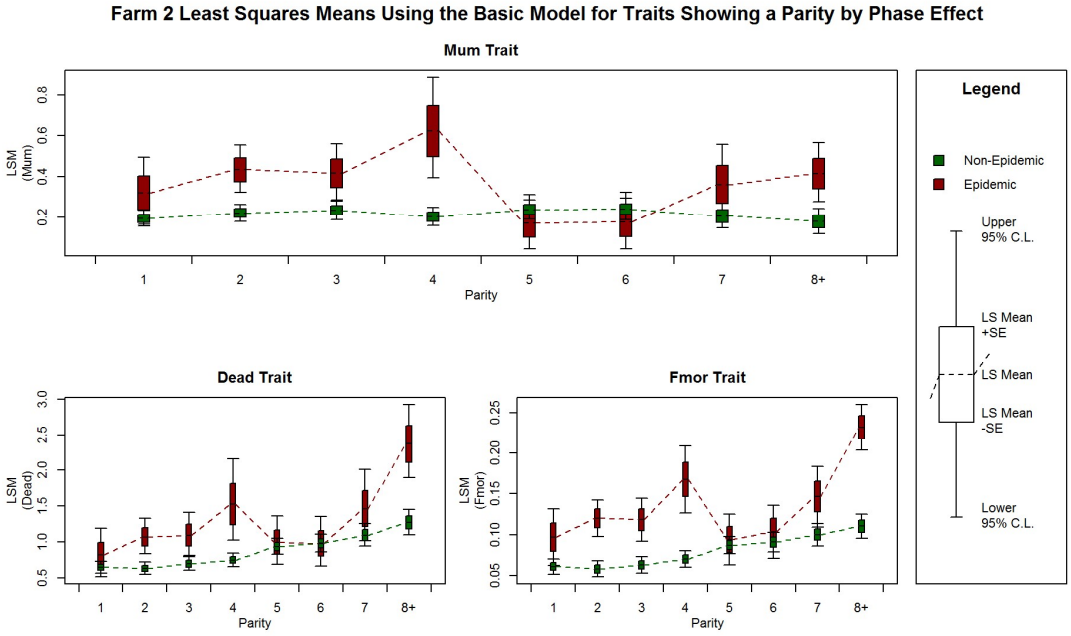

Figure 2.13 – Farm 2 Least Squares Means for All Traits with Parity by Epidemic Effect Estimated Under Model 1

Only traits with a significant Phase ×Parity interaction are shown. *Mum* and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

A systematic parity effect is observed in the absence of a PRRSV outbreak for the *Dead* and *Fmor* traits though not so for the *Mum* trait which appears to remain constant. Similar patterns (with higher numbers) are observed during the epidemic period, though a string increase in *Dead* and *Fmor* LSM was observed in parities greater than seven. All three loss traits show an increase at parity four however the increase in standard error suggests this is more likely an artefact of the data, rather than a systematic effect.

Mixed models were also explored on data including the unconfirmed epidemic. Tables of these results are given in the appendix showing: P values for all fixed effects retained in the model following the stepwise regression process (Table A.2); relative quality of the models across models and traits (Table A.3); and Least Squares means estimates by Phase and Epidemic (Table A.4).

Figure 2.14 shows the least squares means for the *Mum* trait for all models. A significant effect is seen by phase regardless of whether the epidemics are assessed together (as in the basic model) or treated independently (as in the alternative models).

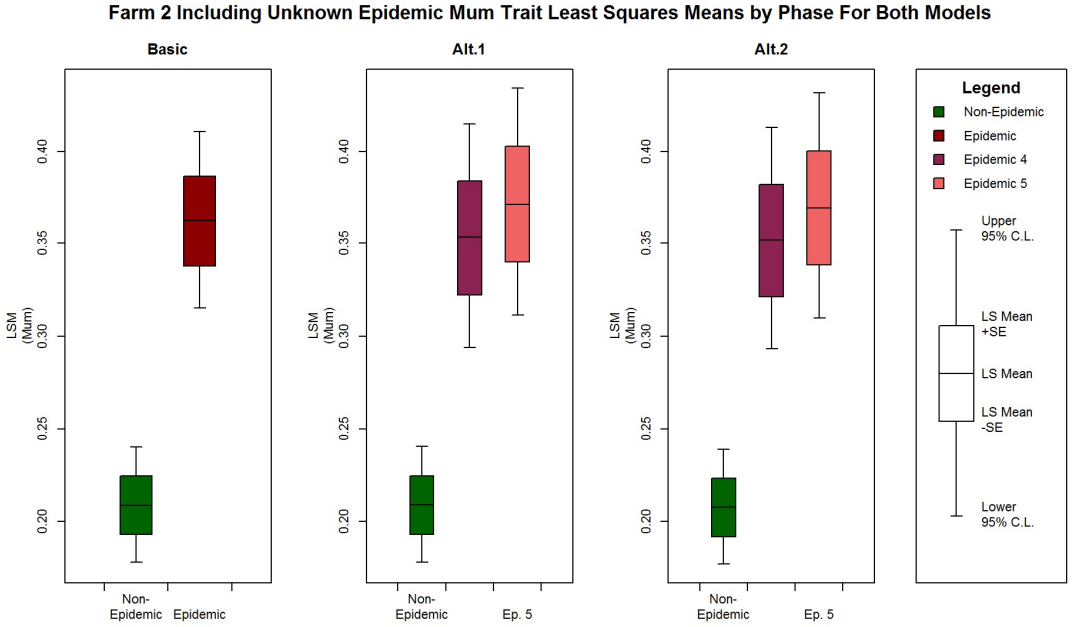

Figure 2.14 – Farm 2 Including Unknown Epidemic Mum Trait Least Squares Means for Epidemic Effect Estimated Using Different Models

Log(*Mum*+1) trait fitted, back-transformed onto the response scale. Covariates involved in interactions with the epidemic factor are estimated at their mean for each epidemic level.

Given the unknown source of epidemic 5 the epidemic differences are considered for all traits under alternative model 1, to consider the two epidemics independently. The least squares means for all traits by Epidemic is shown in Figure 2.15.

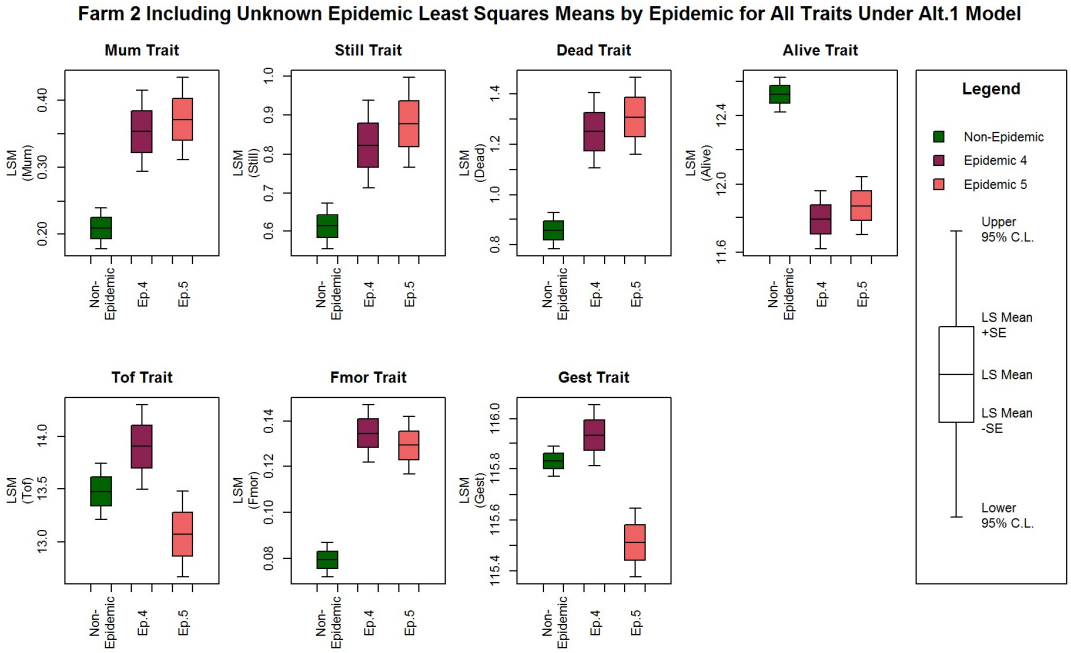

Figure 2.15 – Farm 2 Including Unknown Epidemic Least Squares Means for All Traits by Epidemic Estimated Under Alternative Model 1

Mum, *Still* and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

For all the disease traits identified in the previous analysis (*Mum*, *Still*, *Dead*, *Alive* and *Fmor*) estimates are similar between epidemics but show significant ($p<0.001$) differences between epidemic and non-epidemic phase. This is not the case for the other traits in the analysis. *Tof* and Gest showing no significant difference between phases, though a significant difference is observed between epidemics.

Traits for which a significant epidemic by parity effect was identified were *Mum Dead Alive* and *Fmor*. The least squares means of the epidemic by parity effect for all models for the *Mum* Trait is shown in Figure 2.16.

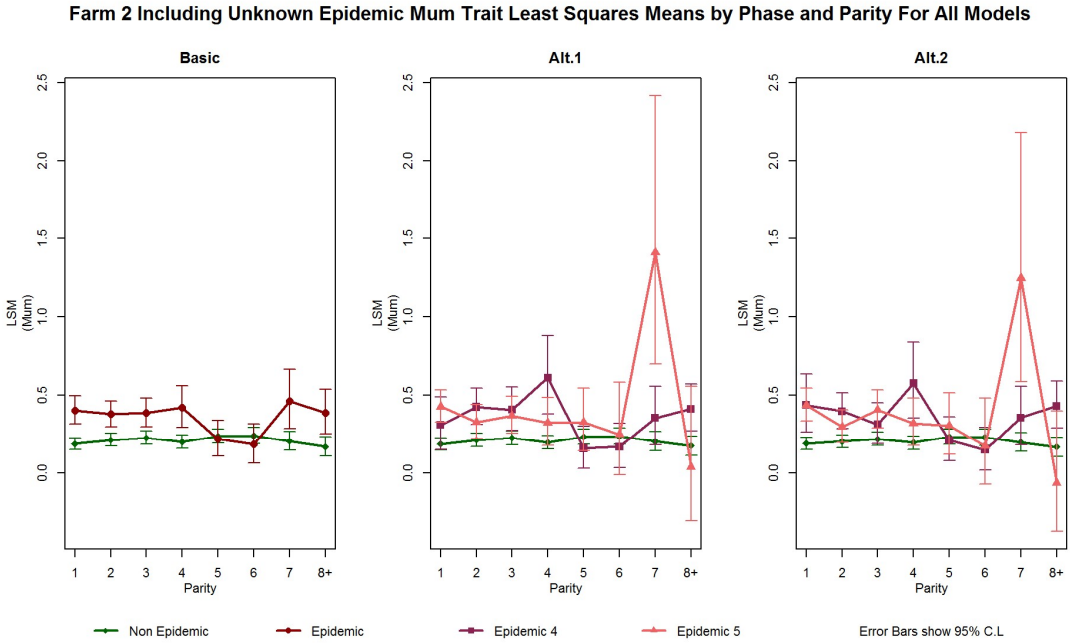

Figure 2.16 – Farm 2 Including Unknown Epidemic Mum Trait Least Squares Means by Parity and Epidemic Estimated Under All Models

Log(*Mum*+1) trait fitted, back-transformed onto the response scale. Covariates involved in interactions with the epidemic factor are estimated at their mean for each epidemic level.

Significant differences are seen for most parity classes between phase under the basic model expect for parities 5 and 6. Similar differences are seen between parity contrasting non-epidemic and epidemic 4 under the alternative models though considerable overlap is seen with epidemic 5 with a much wider confidence interval for some parity classes. This pattern was repeated across all traits for which an epidemic by parity effect could be demonstrated. The epidemic by parity effect for these traits under the basic model is shown in Figure 2.17.

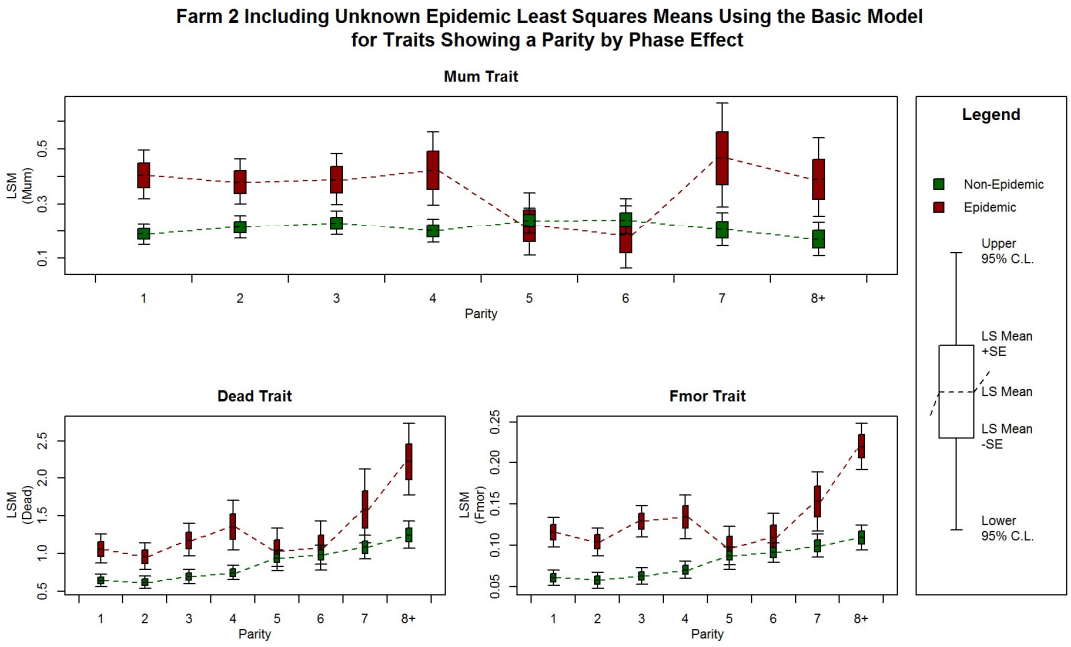

Figure 2.17 – Farm 2 Including Unknown Epidemic Least Squares Means for All Traits With Parity by Epidemic Effect Estimated Under Model 1

Only traits with a significant Phase ×Parity interaction are shown. *Mum* and *Dead*, transformed by log(trait+1), are back-transformed onto the response scale.

No systematic pattern is observed in non-epidemic phase for the *Mum* trait though for *Dead* and *Fmor* a systematic increase is seen in later parities in LSM estimates.

2.04 Discussion

The results of this chapter show a considerable, and consistent decrease in reproductive performance during periods of PRRSV outbreaks. This was found across the *Mum, Still, Dead Alive Fmor* and *Wean* traits, demonstrating their usefulness as disease indicator traits in both datasets.

Both farms demonstrate significant differences in LSM estimates by phase, for traits associated with reproductive losses, a known clinical consequence of PRRSV infection. For traits reported in terms of numbers of piglets, the *Alive* trait shows the greatest contrast on both farms. The differences in LSM between non-epidemic and epidemic phase (the *epidemic impact*) for the *Alive* trait was -2.807 and -0.73, for farm 1 and farm 2 respectively.

The within farm between epidemic difference is considerable for farm 1; epidemic 1 shows the greatest epidemic impact on the *Alive* trait of -3.92; the epidemic with the smallest impact is epidemic 3 of -1.342. Based on the available information it is not possible to determine the underlying causes of the difference in epidemics these could include differences in PRRSV strains, risk of previous exposure or co-infection with other pathogens.

Similar effects are observed as reported in other studies investigating reproductive failure associated with PRRSV (Lewis *et al.,* 2009a, 2009b; Rashidi *et al.,* 2014; Serão *et al.,* 2014). These studies found that parity was a significant factor explaining variation in reproductive outcomes during PRRSV outbreaks. Whilst Serão et al. (2014) and Rashidi et al. (2014) found the effects of contemporary group to be significant this was not observed in this study, as found in Lewis *et al.,* (2009a).

Both previous studies that investigated the effect of sow line on reproductive outcomes found this to be significant (Lewis et al., 2009a; Serão et al., 2014). The main difference between this study and previous studies is the inclusion of litter size (*Tof*) as a fixed covariate in the model. When *Tof* was omitted from the model sow line was consistently significant whereas when included the sow line effect was non-significant.

Previous analyses of the traits considered here did not include *Tof* as covariate in the statistical models. However, PRRS and other factors such as sow line genetics, considered in the statistical models may affect both, the total number of foetuses (*Tof*), as well as the survivability of the foetuses. Hence, exclusion of *Tof* as covariate for reproductive failure traits may confound the effects of the model predictors on *Tof* and the trait under consideration. For this reason, *Tof* was included as covariate in the models of this study. However, it is important to note that this not only affects the trait definition (e.g. number of mummified piglets corrected for *Tof* rather than number of mummified piglets overall) but

may also affect the comparison of the results of this study to those that did not explicitly account for *Tof*.

Consideration of the litter size in model for prenatal and perinatal mortality is also important because of inter-uterine crowding cited as a significant factor in prenatal mortality (Dziuk, 1968); with foetal development impaired as litter size increases (Vonnahme et al., 2002; van der Waaij et al., 2010). As observed in Johnson et al., (1999) litter size also has a significant effect on piglet birth weight It is however acknowledged that the inclusion of *Tof* in the model may obscure any interaction between PRRSV pathogenesis and litter size; *Tof* was included to account for the documented effects of litter size on prenatal (and where applicable preweaning) mortality.

The least squares means under this analysis found slightly greater impacts for most disease traits than observed in Lewis *et al.,* (2009a). The only disease trait which did not show increased impact was *Mum* with a LSM difference of 0.934 compared to a small difference of 1 in the previous study. This is most likely due to differences in the models (e.g. inclusion of *Tof* as a covariate) and a more stringent definition of disease phase pertinent to mortality and survival type traits.

Consideration of exposure plays an important part in the analyses of field data, on which this study was based. Field data is more widely available and considerably cheaper than experimental data. While experimental data offers a more controlled environment for recording results, the number of individuals which may be used and the number of measurements that can be taken is often limited by financial constraints. The study by Ladinig *et al.,* (2014) found that experimental infection at day 85 of gestation resulted in a foetal mortality of 41% (s.e. 22.8%). The foetal mortality observed in the epidemic phase of 33.4% (s.e. 0.01%) for farm 1 is well within the bounds of the experimental estimates though the estimate for farm 2 is slightly lower at 0.13% (s.e. 0.9%).

It is impossible to relate the categorical differences found in this study to the differences observed in Ladinig *et al.,* (2014) which included the traits autolysed and meconium-stained. In the challenge experiment inoculation occurred at gestation day 85, too early for the *Mum* trait to be realised which requires death prior to ~35 days of gestation (Christianson, 1992). The categories such as *Mum* or autolysed take time to develop *in utero*, as such the ratios of different categories are determined by time point at which the foetus dies, impacted by stage of gestation at the time of infection.

Field data may have the adequate high number of animals, but are inherently noisy (Bishop and Woolliams, 2010). It is well known that the host genetic contribution to resilience and resistance traits depends on exposure. In natural disease outbreaks the exposure is usually not known and is likely to vary across and within outbreaks. An inability to account for incomplete or varying exposure can lead to a downward bias of heritability estimates (Bishop & Woolliams, 2010; Bishop *et al.,* 2012). Accounting for disease severity is thus an important consideration especially when utilising field data when the infection status of the individual is unknown.

In this study, differences in exposure across and within different PRRSV outbreaks were explicitly accounted for in the alternative models 1 and 2. Whereas alternative model 1 only accounted for differences between the epidemics (by fitting epidemic id), alternative model 2 also accounted for differences in exposure within the epidemics by fitting a rolling average of reproductive performance traits. Both alternative models not only provided a superior model fit to the data, but also provided relevant new insight into the effect of the severity of the outbreak on the diverse reproductive performance traits. Fitting the trend to account for differences in exposure was introduced during this research, and adopted by Serão *et al.,* (2014). They found that this trend better accounted for contemporary group than estimates using breeding month and farrowing year.

Overall, the results for alternative models 1 and 2 were very similar, indicating that variation in exposure within the individual outbreaks had a significant, though not a strong effect on the least square mean estimates of the reproductive performance traits.

It should be noted that fitting epidemic identifier or the rolling average trait trends in the statistical models could generate potential confounding between the estimates of the random sow effects and the rolling trait averages, if only one record per sow is available in each epidemic. This was indeed the case for all the sows in both, farm 1 and farm 2 datasets. This confounding could prove problematic in generating unbiased heritability estimates using the alternative models.

Parity of the sow was found to have a significant impact on reproductive performance in the absence of and during the PRRSV outbreaks. For farm 1, similar parity profiles in response to disease are seen using the comparable basic models to those presented Lewis *et al.,* (2009a). However, on farm 1 when the epidemic ID was accounted for this pattern was lost. Whilst outbreak was considered as confounding pre-exposure in Lewis *et al.,* (2009a) only two outbreaks were considered in that analysis. In this study results of the parity effect in the individual epidemics may suggest a systematic effect in Epidemic 1 and Epidemic 3. Given a lack of representation of parities in all epidemics; lack of information on individual immune status; and the confounding of age, parity, risk of pre-exposure, and epidemic ID in the data; further research would be required to demonstrate this. The older (higher parity) animals in Epidemic 2 show evidence of a similar effect as that described for epidemic 3 whilst younger (lower parity) animals show evidence of the effects described for epidemic 1. This could occur if older animals were present on the farm during epidemic 1, whereas younger animals were PRRSV naïve. These results may suggest a pattern of higher losses in higher parities in PRRSV naïve animals which is reversed in non-naïve animals. If this is the case, the current practice to challenge animals with PRRSV (using vaccination or live virus) as part of the

acclimation process on health challenged farms, is highly beneficial in mitigating higher losses as a result of infection in later parities.

The “unconfirmed epidemic” in farm 2 is interesting. This outbreak of reproductive losses was identified by a change in the 30 day rolling averages across all traits, the overlapping nature of which could indicate a single source outbreak. However, the pathological basis of this without a positive diagnostic test remains unknown. Given that regular PRRSV testing implemented on the farm had a specificity of 99.6%, the cause for the observed reduction in reproductive performance is unlikely (though not impossible) to be PRRS. The least squares means of the two epidemics on farm 2 were very similar for the disease indicator traits (*Mum, Still Dead, Alive* and *Fmor*) though significant differences were seen in the traits; *Gest* and *Tof.*

In conclusion, the results of this chapter provide quantitative estimates for the effect of PRRSV on a number of reproductive performance traits, and the base models for the subsequent genetic analyses. In addition results indicate a previously unidentified systematic parity effect in naïve animals of increased losses in higher parties and suggest that this pattern maybe reversed in animals with prior exposure.

Chapter 3. Genetic Variation of Sow Reproductive Performance in the Absence of and During PRRSV Outbreaks

3.01 Introduction

The literature review covers a range of studies demonstrating a genetic basis of response in terms of response to PRRSV both in terms of breed effect and heritable component. The heritable component is important in understanding the extent to which selection can be used to improve the response of pigs to PRRSV infection. It is this heritable component that is exploited in selection strategies. In conjunction with genetic data it can be used to look at the distribution of variation across the genomic variants for use in selection methods. Not only do these selection methods mitigate some of the losses associated with disease but regions which show heritable effects can be investigated to advance understanding of the virus-host interaction.

Heritability has been estimated under PRRSV challenge, for a number of traits which quantified response in terms of specific outcomes. While some studies have dissected response using direct measures of disease burden or immunological response (see literature review) several have used reproductive traits comparable to those used in this research. These methods enable the direct quantification of losses in terms of one of the main economic units of pig production (i.e. number of piglets produced).

Previous studies have provided heritability estimates for reproductive performance traits under PRRS outbreaks (Lewis et al., 2009a and Serao et al., 2014). As already pointed out before, there is a difference in the models derived in Chapter 2 with regards to those presented in both Lewis et al., (2009a) and Serao et al., (2014) in that the models presented here include *Tof* as a fixed covariate. The heritability of the *Still* trait is generally considered low when the trait is corrected for litter size (reviewed in Knol et al., (2002)) whist the heritability of litter size itself is higher averaging ~0.10 in the literature (Rothschild & Ruvinsky, 2011). Given litter size is shown to have a positive genetic correlation with number of stillborn piglets this direct effect may upwardly bias estimates of the heritability of the number of stillborn piglets when investigating prenatal and perinatal survival when the

total litter size is not accounted for. Heritability estimates of prenatal mortality (*Mum, Still* and *Dead*) are reported higher when litter size is not accounted for at ~0.17 (i.e. Johnson et al., (1999)).

Lower heritability estimates under normal environmental conditions (i.e. the absence of PRRS) may allow for greater contrast of the heritability of mortality in the presence of PRRS. Such a contrast could therefore provide a useful genetic signal for exploring additive genetic effects attributable to resilience with respect to survival and mortality traits.

The variation of reproductive response to PRRSV as explored in Rashidi et al. (2014) demonstrates similar models to those used in the previous chapter can be used to estimate both the effect on the herd in terms of disease status and effect of the individual in terms of variation in response. Lacking in pedigree or genetic data however, the effect of the individual in the repeatability model (fitted as sow identity) confounds both genetic and permanent environmental effects.

The heritable component of reproductive performance, is explored in Lewis *et al.,* (2009b), using similar methods as employed in this study, for farm 1 data described in section 2.02.1. Pedigree information was used to estimate genetic parameters for reproductive traits using a model similar to the basic model described in the previous chapter, the differences highlighted in the Discussion. The Lewis *et al.,* (2009b) study reported low/moderate heritabilities (0.12-0.22) in epidemic phase. For disease indicator traits (*Mum, Still, Dead, Alive, Wean*) an increase in heritability estimate is observed in epidemic phase as compared to non-epidemic phase as a result of a proportionally greater increase in additive genetic variance as compared to environmental variance. Here new Illumina platform 60K SNP data is available to use direct genetic data to quantify additive genetic variance for this farm. In addition, an independent dataset (farm 2) is available in this study.

Serão *et al.,* (2014) estimated genetic parameters for reproductive performance (in addition to an analysis of sample to positive ratio of antibody response) in PRRSV epidemic phase with 525 litter records. The mixed model used accounts for non-genetic sources of variation including the rolling 30-day trait average to, account for the dynamic change in severity over the time course of the outbreak, equivalent to alternative model 2 described in the previous chapter. In addition Serão *et al.,* (2014) used the *Fmor* trait, used in the previous chapter described in Orrett *et al.,* (2013). Following this analysis low to moderate heritabilities (0.1-0.12) were reported for the reproductive traits analysed, with many traits showing an increase in heritability in epidemic phase as compared to non-epidemic phase. Some traits were conflicted by low estimates and high standard errors suggesting issues with statistical power.

As outlined in the general introduction BLUP is the most widely used tool for estimating additive genetic variance (σ^2_A). This process fits a relationships matrix describing the pairwise additive relationships within the population. These relationship matrices can take several forms but include the additive relationship (A) matrix and a genomic relationship matrix (G or GRM). A matrices are calculated using the ancestry defined in a pedigree to estimate kinship. G Matrices are calculated comparing genetic data between individuals to estimate kinship. These values are used to represent the probability that two alleles at a given locus are not just identical by state (IBS), but a result of identity by descent (IBD).

The A matrix captures only the discrete family relationships detailed in the pedigree and assumes that founders are unrelated. This limits the relationships in the matrix to those animals for which ancestry is captured in the pedigree. G matrices, in comparing genotypes identical-by-state may be more accurate in being able to assess relationships where the ancestry is unknown, but can be sensitive to changes in allele frequency in the observed population when estimating relationships over a more ancient, unknown, ancestry (Powell *et*

al., 2010). While the G matrix is therefore more accurate, the assumptions used in A matrix production can make it more robust to population stratification (Amin *et al.,* 2007).

This chapter uses pedigree (where available) and genomic information in conjunction with models identified in the previous chapter to estimate variance components and heritability for reproductive performance of sows during and in the absence of PRRSV outbreaks. Some success has been reported in the literature of combining data from different farms to expand the sample size and improve power (Gredler *et al.,* 2007; Riggio *et al.,* 2014). For this purpose, a joint analysis; combining the data between farms by phase, will also be conducted to assess the degree of heritability which may be estimated across all animals available. The effect of incorporating the unknown epidemic described in section 2.03.1 (Farm 2) will also be considered.

Demonstrating that resilience of sows to PRRSV, represented by high reproductive performance under PRRSV challenge is heritable, is the first step to determine whether breeding for resilience is feasible. It may also be used to explore the genomic architecture underlying variation of response, in subsequent chapters.

3.02 Materials and Methods

3.02.1 Data

The data used in this chapter is described for farm 1 and for farm 2 in section 2.02.1. Data were partitioned into epidemic and non-epidemic phases according to the adapted threshold-threshold method as described in section 2.02.2.

Some of the subsequent statistical models were unable to account for repeated records per animal. As such, for each analysis, a single record dataset was created containing only one record per animal. Where multiple records per animal were available within a phase only one was selected (i.e. within ‘Farm 1 epidemic phase dataset’ only one litter record per animal was selected in the single record form ‘Farm 1 epidemic phase single record dataset’).

Several methods were considered for which record to select (closest record to the peak of the epidemic identified in the trend analysis, record with the maximum or minimum number of dead piglets etc.). Given a risk of systematic bias occurring from such selection methods the 'first record per animal' in each phase was chosen to mitigate against such biases. Whilst this may have favoured earlier parity animals in the single record datasets, given that earlier parities are better represented in the data, the parity specific effect is likely to be estimated in the model with a smaller error. These heritability estimates are obtained for use in subsequent association analyses.

On both farms, for a subset of animals, SNP genotypes were available generated using the Illumina PorcineSNP60 chip (Ramos *et al.,* 2009) see also (Illumina, 2015). For analyses using this genetic data the phenotype data was limited to the animals that had genotype data and which passed genetic quality control (QC).

A joint analysis was also run whereby the data from the two farms was combined by phase (i.e. non-epidemic/epidemic) to look at potential improvements in power as compared to the individual farm analyses.

Farm 1

Of the 6,369 records for Farm 1 used in the previous chapter there were 5,307 non-epidemic phase records and 1,062 epidemic phase records in the partitioned datasets. Given the reduction in the number of records in the epidemic phase, a considerable reduction in the number of repeated records was seen in epidemic phase with only 71 sows having repeated records. There was insufficient time for a sow to farrow, return to oestrus, conceive and farrow again within the same epidemic, as such these repeated records were distributed across different epidemics. 1,053 sows had repeated records in the non-epidemic phase.

Pedigree Information

A full pedigree was available for farm 1 covering a maximum of 12 generations, visualised in Figure 3.1. In total, the pedigree contained 4,249 animals, 2,691 parents of which 810 were sires and 1,881 were dams.

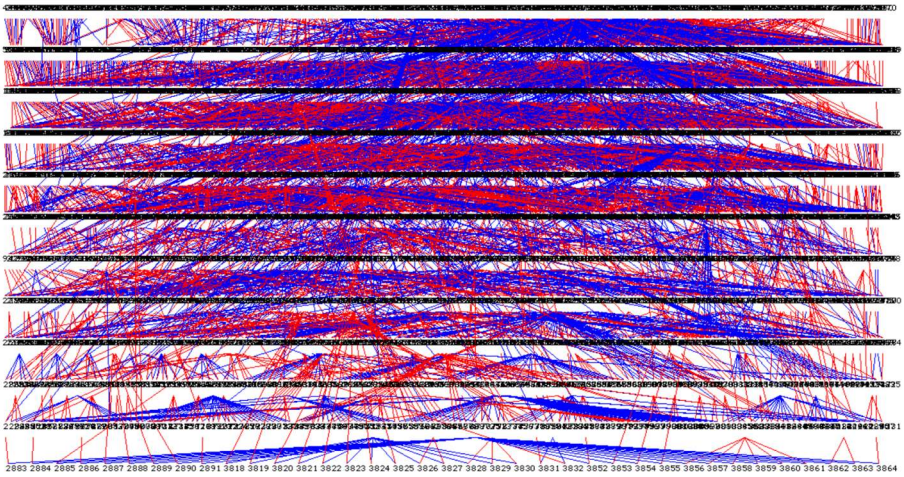

Figure 3.1 – Pedigree Visualisation for Farm 1
A visualisation of the pedigree, connecting sow (red) and sires (blue) to their progeny.

This pedigree is large and complex and highlights an amount of inbreeding as shown by branches spanning multiple generations creating loops in the pedigree.

Genetic Data

For farm 1 SNP data was available for 960 animals and for 60,674 SNPs. Quality control (QC) was run on the available SNP data. Of the 60,674 SNPs 3,534 had a call rate of less than 90% and were excluded, 8,990 had low (<5%) minor allele frequency and were excluded, of which 9 SNPs fell into both exclusion categories, leaving 48,141 SNPs passing all QC criteria. Of the 960 animals, 41 animals were excluded because of low (<90%) call rate, 3 animals excluded because too high autosomal heterozygosity (FDR<1%), 6 were excluded because of assumed errors in labelling and/or tracking of samples (corrected Identity-by-State (IBS) value > 1). In total 910 animals passed the QC criteria.

For analyses using the genetic data the phenotype data was limited to the animals that had genotypic data and passed genetic QC. Limiting the phenotypic data to just the sows for which genetic data was available and passed QC resulted in 707 sows with 2,219 litters in the non-epidemic phase and 605 sows with 651 litters in the epidemic phase.

Farm 2

Of the 5,725 records for Farm 2 used in the previous chapter there were 5,386 non-epidemic phase records and 339 epidemic phase records in the two partitioned datasets. In Epidemic data including the unknown epidemic this increased to 683. Given there was insufficient time for a sow to farrow, return to oestrus, conceive and farrow within the same epidemic, no repeated records were observed in the epidemic phase data. When the unknown epidemic was included 56 sows had records in data covering both epidemic periods.

Genetic Data

For farm 2, no pedigree information was available. However, SNP data was available for 593 animals and 57,786 SNPs. The same for quality control criteria was applied to these data as was applied to farm 1 SNP data. Of the 57,786 SNPs, 476 had a low call rate (<90%), 9,574 had low (<5%) minor allele frequency and 56 SNPs fell into both exclusion categories, leaving 47,680 SNPs passing all quality control criteria. Of the 593 animals, 3 animals were excluded because of low (<90%) call rate, 4 animals excluded because too high autosomal heterozygosity (FDR<1%), 1 animal was included in both these exclusion categories, leaving 586 animals passing all quality control criteria.

Limiting the phenotype data to animals for which genetic data was retained resulted in 585 sows with 2168 litters in the non-epidemic phase and 276 sows with an equal number of litters in epidemic phase, which increased to and 329 sows with 308 litters when the unknown epidemic was included.

Joint Analysis

Genetic Data

QC was performed on the combined farm raw genotype data to account for subtle changes in MAF and call rate. SNPs were included where present in both datasets to avoid confounding by farm leading to 57,440 SNPs in the joint analysis for 1,553 animals. Of the 57,400 SNPs, 1021 had a low call rate (<90%,) 8,258 had low minor allele frequency (<5%), an additional 148 SNPs fell into both exclusion categories, leaving 48,013 SNPs passing all quality control criteria. Of the 1,553 animals, 47 animals were excluded because of low (<90%) call rate, 1 animals excluded because too high autosomal heterozygosity (FDR<1%), 5 animals fell into both these exclusion categories and 6 were excluded because of assumed errors in labelling and/or tracking of samples (corrected Identity-by-State (IBS) value > 1) in total 1,495 animals passed quality control criteria.

3.02.2 Constructing Relationship Matrices

The A matrix was calculated by ASReml (Gilmour *et al.,* 2008) according to the method described by (Meuwissen & Luo, 1992). This calculates relationships based on an equal contribution of each parent for each offspring, using a known pedigree, correcting for inbreeding.

The GenABEL package in R (Aulchenko *et al.,* 2007b) was used to run quality control on the genomic data, generate principal components and produce the genomic relationship matrices (GRMs) in accordance with the specification in (Gilmour *et al.,* 2008).

The G matrix was calculated as described in (Uemoto *et al.,* 2013) based on (Aulchenko *et al.,* 2007b). For between distinct individuals ($i \neq j$) IBS values this is shown in Equation 3.1 for self-similarity ($i=j$) this is shown in Equation 3.2.

$$f_{ij} = \frac{1}{n}\sum_{k=1}^{n}\frac{(x_{ik} - p_k) - (x_{jk} - p_k)}{p_k(1 - p_k)}, (i \neq j)$$

Equation 3.1 – Pairwise Identity-by-State calculation between individuals

Where the f_{ij} is the identity by state (IBS) between individuals *i* and *j*; x_{ik} and x_{jk} is the genotype of SNP *k* (of *n* SNPs) for the *i*th and *j*th individuals respectively, coded as: 0, for AA; 0.5 for AB and; 1 for BB, and p_k is the frequency of the A allele.

$$f_{ij} = 1 + \frac{1}{n}\sum_{k=1}^{n}\frac{Obs(\#Hom)_{ik} - Exp(\#Hom)_k}{1 - Exp(\#Hom)_K}, (i = j)$$

Equation 3.2 – Self-similarity Identity-by-state calculation

Where $Obs(\#Hom)_{ik}$ and $Exp(\#Hom)_{ik}$ is the observed and expected number of homozygous genotypes (under HWE) at SNP *k* (of *n* SNPs) for individual *i, respectively*.

The inclusion of for allele frequency (p_k) in Equation 3.1 provides a weighting for allele frequency and in Equation 3.2a correction is made for observed - expected homozygosity (based on allele frequency). This can result in smaller than expected values between unrelated individuals (i.e. < 0) and higher than expected self-similarity values (i.e. >1). Similarly, the correction made in the A matrix calculation, to correct for inbreeding can lead to values >1. This should be taken into account when looking at the plots of the relationship matrices.

Twice the IBS was used to calculate an estimate of IBD for the Genomic Relationship Matrix (Amin *et al.,* 2007; Aulchenko *et al.,* 2007b).

3.02.3 Statistical Analyses

Response traits included in the analysis (*Mum, Still, Dead, Alive, Tof, Fmor, Wean* and *Gest*) are described in section 2.02.1 summarised in the appendix in Table A.1. For the Farm 2 and

joint analysis *Wean* was dropped as a trait due to an inability to account for fostering decisions on farm 2. The traits *Mum, Still and Dead* were transformed log(*Trait*+1).

Additive genetic variance was estimated fitting a linear mixed model using Restricted Maximum Likelihood (REML) (Patterson & Thompson, 1971) using the ASReml 3.0 package (Gilmour *et al.,* 2009). The general linear mixed model equation is shown in Equation 2.1. Using an animal model the additive genetic variance (σ^2_A) is estimated fitting a random effect distributed $\sim N(0, R \cdot \sigma^2_A)$. Where *R* is the relationship matrix calculated using pedigree information (A Matrix) or genomic information (G matrix). In this method both fixed effects and the genetic merit of the individuals within the population can be derived using the Best Linear Unbiased Prediction (BLUP) (Henderson, 1950). This "genetic merit" estimated using the A matrix are termed Estimated Breeding Values (EBVs) or when fitted using the G matrix, Genomic Estimated Breeding Values (GEBVs).

The fixed and non-genetic random effects included in the models for the partitioned data analysis mirror those used in the previous chapter. To determine the relevant fixed and random effects for the models in the partitioned data the same stepwise backward elimination method described in 2.02.3 was used to find the most comprehensive, but conservative (in terms of over-fitting) model pertinent to the trait under evaluation. To ensure that models were hierarchically well formulated lower order terms are retained (even where not significant) when involved in higher order interactions (Kleinbaum *et al.,* 2008). Sow line was retained in all models to account for breed effects, regardless of AIC or F-test p-value given the considerable breed structure within Farm 1 and the significant Sow Line effect demonstrated in the previous chapter for Farm 2. For the joint analysis, the Farm identifier (1 or 2) was added as an additional fixed effect.

Following the stepwise elimination method, the final basic models fit sow line and parity in all models, *Tof* was included as a covariate for the traits. *Mum, Still Dead, Alive Gest,* and (where included) *Wean*. In farm 1 the sow line × parity interaction was additionally fitted for

Alive and *Tof* traits and the net fostered covariate and *Tof* × net fostered covariate interaction fitted for the *Wean* trait. Additionally, in the joint analysis the farm fixed effect was significant and thus retained in all models. The 'phase' term (in the basic model in chapter 2) is dropped given the phases are to be assessed separately, providing variance components and heritability estimates for each phase (Epidemic/Non Epidemic/Epidemic Including Unknown) independently.

Given the considerable population structure suggested by the breed diversity in farm 1 principal components were also considered as a means of accounting for stratification, parallel to fitting sow line. The first 10 principal components were investigated for their ability to account produce K-mean clustering (Liu *et al.,* 2013) mirroring known sow line information. Due to the presence of crosses in the data sow lines do not appear as discrete clusters and so principal components were included as covariates (Price *et al.,* 2006). The first three principal components were included. These were applied to the statistical models for both the Farm 1 and Joint analysis though not for Farm 2 (see results).

To account for differences between the discrete epidemic periods alternative model 1 (Alt.1) additionally fits a unique identifier allocated to each epidemic in the partitioning process (see 2.02.2). In the joint analysis given the hierarchical structure Epidemic ID was nested within Farm, while still including a separate Farm specific effect.

To account for the dynamic trend within each epidemic, in alternative model 2 (Alt.2) a trait trend is fitted as a fixed covariate for each epidemic, additional to the Alt.1 terms, as outlined in the previous chapter. Similarly, in the joint analysis, this effect was nested within Farm, while still including a separate Farm specific effect.

To account for repeated records (in analyses using multiple record per sow data) a repeatability model was included to account for the permanent environmental effect (σ^2_{PE}); (Lush, 1937). This accounts for the non-genetic effects associated with the shared common

environment among littermates, a potential source of bias in estimating additive genetic variance. This effect fits the sow identifier as an additional random effect in the model.

The significance of fixed effects and covariates was assessed under the conditional Wald F-Statistic and corresponding p-values. The significance of random effects was calculated under the Likelihood Ratio Test (LRT) statistic taken following an equal mixture of the χ_0^2 and χ_1^2 distributions (Self & Liang, 1987). The LRT was calculated as twice the difference between the maximum log likelihood of the model excluding the random term and that of the model with its inclusion. While unsuitable for the formal testing of significance the ratio of additive genetic variance to its standard error was used as an indicator of power.

In the first instance, the additive genetic variance was estimated for all traits and models on the individual farms by phase (Epidemic, Non-Epidemic and where applicable Epidemic Including Unknown). Subsequently the combined farm data was analysed using all traits and models to estimate the additive genetic variance in the joint dataset. Finally, to consider the impact on power of the various models; the incorporation of the unknown epidemic in the analysis; and the combining of the datasets across farms the additive genetic variance, standard error and the ratio of the two was also given consideration.

Additional packages used in the visualisation of data included gdata (Warnes *et al.,* 2015), plot3d (Soetaert, 2016), pedigreemm (Vazquez *et al.,* 2010) and Spatstat (Baddeley *et al.,* 2015)

Phenotypic variance (σ^2_P) was calculated as the sum of the environmental (residual) variance (σ^2_E) plus the sum of other variance structures in the model (σ^2_A and, where fitted, σ^2_{PE}). Narrow sense heritability (h^2) is given as the additive genetic variance component divided by the total phenotypic variance (Falconer & Mackay, 1996).

3.03 Results

3.03.1 Farm 1

Summary statistics for all datasets are presented in Table 2.4. All animals used in the analysis in chapter 1 had pedigree records available for A matrix calculation and subsequent variance component estimation. Trait means for these are shown in Table 2.4.a. The single record dataset selected the first record per sow in each phase for which marginal differences are seen in the trait means (Table 2.4.b). Given earlier epidemics were demonstrated to have higher impacts, the trait means in the partitioned data can be seen to be slightly higher. For the analysis of genotyped animals passing QC, considerably fewer numbers of records are available, which also affected the trait means and SEMs (Table 2.4 c). This was also the case for the single record dataset for genotyped animals shown in Table 2.4.d.

Table 3.1 – Summary statistics showing Counts and Trait Means for Farm 1 Datasets

Level	Records	Sows	*Mum*	*Still*	*Dead*	*Alive*	*Tof*	*Fmor*	*Wean*	*Gest*
a. All Animals Multiple Records										
Non-epidemic Phase	5307	1492	0.23 (9×10^{-3})	0.59 (0.01)	0.82 (0.02)	10.27 (0.04)	11.09 (0.04)	0.07 (2×10^{-3})	9.6 (0.03)	115.6 (0.02)
Epidemic Phase	1062	991	2.18 (0.1)	1.42 (0.07)	3.6 (0.12)	7.19 (0.13)	10.79 (0.1)	0.33 (0.01)	6.77 (0.1)	115.52 (0.07)
Epidemic 1	199	199	3.46 (0.31)	1.03 (0.11)	4.49 (0.34)	5.51 (0.28)	10.01 (0.25)	0.4 (0.03)	5.03 (0.24)	115.23 (0.18)
Epidemic 2	705	705	1.95 (0.11)	1.68 (0.09)	3.64 (0.14)	7.18 (0.16)	10.82 (0.12)	0.34 (0.01)	6.71 (0.12)	115.72 (0.09)
Epidemic 3	158	158	1.57 (0.24)	0.72 (0.11)	2.28 (0.27)	9.35 (0.34)	11.64 (0.24)	0.2 (0.02)	9.23 (0.12)	114.99 (0.17)
b. All Animals Single Records										
Non-epidemic Phase	1492	1492	0.23 (0.02)	0.43 (0.02)	0.66 (0.03)	9.9 (0.08)	10.56 (0.08)	0.07 (3×10^{-3})	9.73 (0.05)	115.65 (0.04)
Epidemic Phase	991	991	2.26 (0.11)	1.43 (0.07)	3.7 (0.13)	7 (0.14)	10.7 (0.1)	0.34 (0.01)	6.69 (0.11)	115.56 (0.07)
Epidemic 1	199	199	3.46 (0.31)	1.03 (0.11)	4.49 (0.34)	5.51 (0.28)	10.01 (0.25)	0.4 (0.03)	5.03 (0.24)	115.23 (0.18)
Epidemic 2	671	671	2 (0.12)	1.67 (0.09)	3.67 (0.15)	7.11 (0.17)	10.78 (0.12)	0.35 (0.01)	6.7 (0.13)	115.75 (0.09)
Epidemic 3	121	121	1.78 (0.3)	0.77 (0.13)	2.55 (0.34)	8.84 (0.4)	11.39 (0.27)	0.23 (0.03)	9.34 (0.13)	115.01 (0.2)
c. Genotyped† Animals Multiple Records										
Non-epidemic Phase	2219	707	0.23 (0.01)	0.55 (0.02)	0.78 (0.02)	10.23 (0.06)	11 (0.06)	0.07 (2e-03)	9.69 (0.04)	115.56 (0.03)
Epidemic Phase Total	651	605	2.26 (0.13)	1.46 (0.09)	3.73 (0.15)	7.12 (0.17)	10.85 (0.12)	0.35 (0.01)	6.96 (0.13)	115.55 (0.09)
Epidemic 1	68	68	3.44 (0.49)	1.01 (0.16)	4.46 (0.53)	5.29 (0.43)	9.75 (0.37)	0.42 (0.04)	5.49 (0.36)	115.07 (0.33)
Epidemic 2	467	467	2.25 (0.14)	1.72 (0.11)	3.96 (0.18)	6.83 (0.2)	10.8 (0.14)	0.37 (0.02)	6.61 (0.15)	115.74 (0.11)
Epidemic 3	116	116	1.63 (0.3)	0.71 (0.13)	2.34 (0.33)	9.37 (0.39)	11.71 (0.27)	0.2 (0.03)	9.23 (0.14)	115.03 (0.19)
d. Genotyped† Animal Single Records										
Non-epidemic Phase	707	707	0.24 (0.03)	0.46 (0.03)	0.7 (0.05)	9.89 (0.11)	10.6 (0.11)	0.07 (5×10^{-3})	9.84 (0.06)	115.64 (0.06)
Epidemic Phase	605	605	2.37 (0.14)	1.49 (0.09)	3.85 (0.16)	6.91 (0.18)	10.76 (0.12)	0.36 (0.01)	6.89 (0.13)	115.59 (0.1)
Epidemic 1	68	68	3.44 (0.49)	1.01 (0.16)	4.46 (0.53)	5.29 (0.43)	9.75 (0.37)	0.42 (0.04)	5.49 (0.36)	115.07 (0.33)
Epidemic 2	451	451	2.29 (0.15)	1.7 (0.11)	3.98 (0.19)	6.78 (0.21)	10.77 (0.14)	0.38 (0.02)	6.63 (0.16)	115.78 (0.11)
Epidemic 3	86	86	1.93 (0.38)	0.76 (0.16)	2.69 (0.42)	8.84 (0.47)	11.52 (0.3)	0.23 (0.04)	9.36 (0.15)	114.98 (0.24)

†Genotyped and passing genetic quality control. Summary statistics for data under different partitions showing number of records, number of sows and means with standard error of the mean for the reproductive performance traits considered in this study

A-Matrix

A visualisation of the additive relationship matrix generated from the pedigree shown as a heat plot in Figure 3.2.

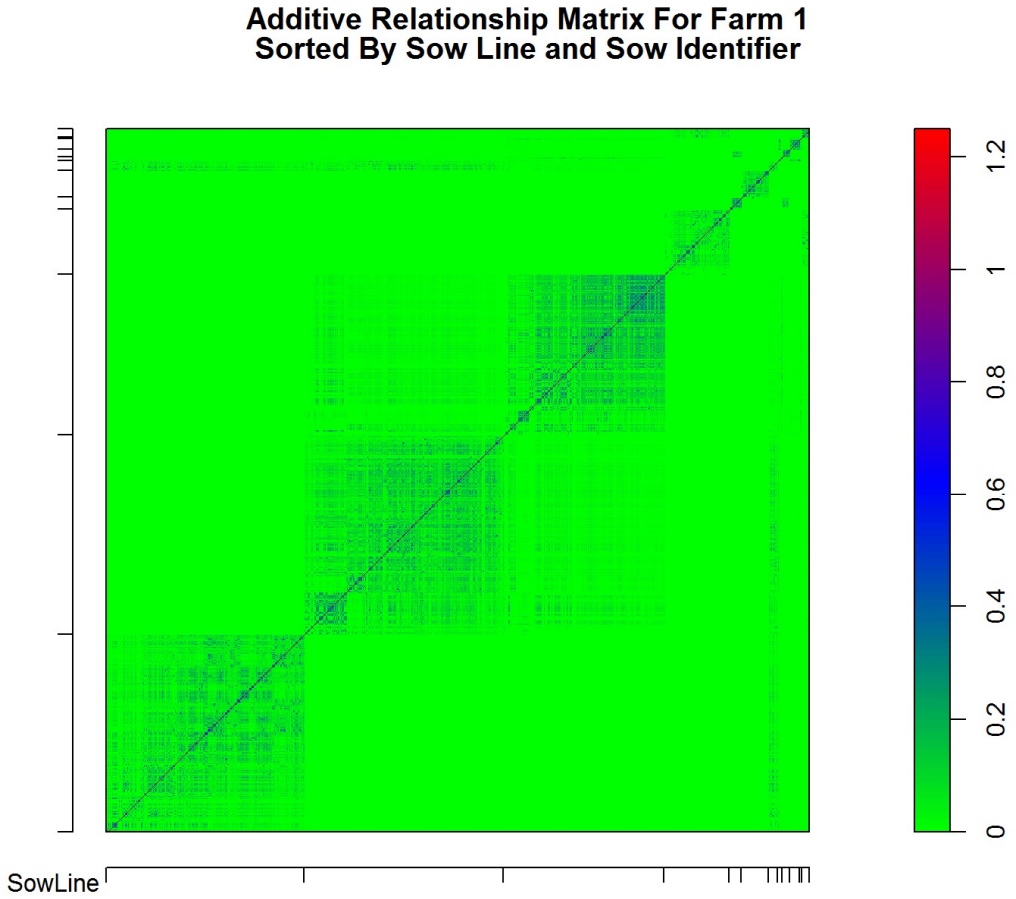

Figure 3.2 – Heat Plot of the Additive Relationship Matrix For Farm 1
Additive relationships showing all sows ordered by sow line and sow identifier (x and y axis).

As can be seen from the range of the key, relationships extend above 1 as a result of inbreeding, increasing self-identity by factor *F,* the inbreeding coefficient. Within line relatedness shows a high degree of close familial type relationships of ~0.5 (blue) i.e. siblings. Between line relatedness was zero for the most part, except for known sow line crosses, which are represented by the off-diagonal blocks of non-zero additive relationships

Variance Component Estimates

For the repeated measures analysis, the additive genetic variance to standard error ratios were (generally) higher than for the single record analysis, with the repeated measures analysis also showing lower standard errors for the heritability estimate. Marginal differences were seen in actual heritability estimate (<0.05). As such results are presented for the repeated measures analysis.

As expected, estimation of a permanent environmental effect was conflicted by low numbers of repeated records for the epidemic phase. For estimates generated using the A matrix for non-epidemic phase data, significant (p<-0.05) permanent environmental effects could be estimated for the traits *Tof, Fmor* and *Gest* under the basic model. During the epidemic phase, significant permanent environmental effects could only be obtained for *Dead, Alive* and *Fmor* traits.

Table 3.2 shows the variance component estimates, heritability and significance of the additive genetic variance of the diverse reproductive traits for both, the epidemic and non-epidemic phases calculated using the A matrix under the repeated measures model. Summary statistics for this dataset are shown in Table 2.4.a.

Results showing variance component estimates, heritability and LRT p-values for random effects, using the single record data can be found in the appendix Table A.5. Summary statistics for this dataset are shown in Table 2.4.b.

Table 3.2 – Farm 1 Variance Components and Heritability Estimated Using the A Matrix with a Repeated Measures Model for Non-Epidemic (Green) and Epidemic (Red) Phase

		Non-Epidemic Phase							Epidemic Phase						
Trait	Model	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum[¥]	Basic	2×10^{-3} (1×10^{-3})	1×10^{-3} (1×10^{-3})	0.1 (2×10^{-3})	0.1 (2×10^{-3})	0.02 (0.01)	<0.001	0.14	0.1 (0.05)	0.1 (0.08)	0.49 (0.08)	0.69 (0.03)	0.15 (0.06)	<0.001	0.13
Still[¥]	Basic	0.01 (3×10^{-3})	1×10^{-3} (3×10^{-3})	0.19 (4×10^{-3})	0.2 (4×10^{-3})	0.04 (0.01)	<0.001	0.35	0.07 (0.03)	0.03 (0.06)	0.33 (0.05)	0.43 (0.02)	0.16 (0.07)	<0.001	0.4
Dead[¥]	Basic	0.01 (3×10^{-3})	3×10^{-3} (4×10^{-3})	0.24 (0.01)	0.25 (5×10^{-3})	0.04 (0.01)	<0.001	0.25	0.1 (0.05)	0.2 (0.08)	0.41 (0.07)	0.71 (0.03)	0.14 (0.07)	0.01	0.01
Alive	Basic	0.05 (0.02)	0.02 (0.02)	1.51 (0.03)	1.59 (0.03)	0.03 (0.01)	<0.001	0.21	2.19 (0.94)	5.08 (1.39)	6.18 (1.1)	13.45 (0.64)	0.16 (0.07)	<0.001	<0.001
Tof	Basic	1.15 (0.24)	0.43 (0.19)	6.45 (0.14)	8.03 (0.18)	0.14 (0.03)	<0.001	0.01	1.64 (0.67)	1.35 (1.1)	6.8 (1)	9.79 (0.46)	0.17 (0.07)	<0.001	0.1
Fmor	Basic	6×10^{-4} (2×10^{-4})	4×10^{-4} (2×10^{-4})	0.01 (3×10^{-4})	0.01 (3×10^{-4})	0.04 (0.01)	<0.001	0.04	0.02 (0.01)	0.05 (0.01)	0.05 (0.01)	0.12 (0.01)	0.14 (0.06)	0.01	<0.001
Wean	Basic	0.08 (0.03)	7×10^{-7} (1×10^{-8})	2.74 (0.06)	2.82 (0.06)	0.03 (0.01)	<0.001	1	4.06 (0.88)	0.3 (1.07)	5.57 (0.9)	9.94 (0.5)	0.41 (0.08)	<0.001	0.54
Gest	Basic	0.93 (0.13)	0.22 (0.08)	1.4 (0.03)	2.56 (0.08)	0.36 (0.04)	<0.001	<0.001	1.06 (0.33)	3×10^{-6} (2×10^{-7})	3.99 (0.3)	5.05 (0.23)	0.21 (0.06)	<0.001	1
Mum[¥]	Alt.1[†]								0.05 (0.04)	0.06 (0.08)	0.55 (0.08)	0.66 (0.03)	0.08 (0.06)	0.03	0.28
Still[¥]	Alt.1[†]								0.04 (0.02)	3×10^{-7} (2×10^{-8})	0.37 (0.02)	0.41 (0.02)	0.09 (0.05)	0.03	0.89
Dead[¥]	Alt.1[†]								0.04 (0.04)	0.1 (0.09)	0.53 (0.09)	0.67 (0.03)	0.05 (0.05)	0.14	0.18
Alive	Alt.1[†]								1.05 (0.76)	4.2 (1.53)	7.51 (1.37)	12.76 (0.59)	0.08 (0.06)	0.04	0.01
Tof	Alt.1[†]								0.61 (0.5)	1.6 (1.1)	7.05 (1.04)	9.26 (0.42)	0.07 (0.05)	0.06	0.06
Fmor	Alt.1[†]								0.01 (0.01)	0.04 (0.01)	0.06 (0.01)	0.11 (0.01)	0.08 (0.06)	0.07	0.01
Wean	Alt.1[†]								0.6 (0.45)	6×10^{-7} (4×10^{-8})	7.65 (0.51)	8.25 (0.37)	0.07 (0.05)	0.09	1
Gest	Alt.1[†]								1.01 (0.36)	0.17 (0.71)	3.82 (0.66)	5.01 (0.23)	0.2 (0.07)	<0.001	0.59

		Non-Epidemic Phase							Epidemic Phase						
Trait	Model	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum¥	Alt.2	2×10^{-3} (1×10^{-3})	1×10^{-3} (1×10^{-3})	0.1 (2×10^{-3})	0.1 (2×10^{-3})	0.02 (0.01)	<0.001	0.12	0.02 (0.03)	0.05 (0.07)	0.47 (0.07)	0.55 (0.02)	0.04 (0.05)	0.16	0.28
Still¥	Alt.2	0.01 (2×10^{-3})	2×10^{-3} (3×10^{-3})	0.19 (4×10^{-3})	0.19 (4×10^{-3})	0.03 (0.01)	<0.001	0.19	-	0.08 (0.04)	0.26 (0.04)	0.34 (0.01)	-	-	0.05
Dead¥	Alt.2	0.01 (3×10^{-3})	3×10^{-3} (4×10^{-3})	0.24 (0.01)	0.25 (5×10^{-3})	0.03 (0.01)	<0.001	0.21	-	2×10^{-3} (0.08)	0.52 (0.08)	0.52 (0.02)	-	-	1
Alive	Alt.2	0.05 (0.02)	0.02 (0.02)	1.5 (0.03)	1.57 (0.03)	0.03 (0.01)	<0.001	0.19	4×10^{-3} (0.42)	1.8 (1.47)	8.25 (1.45)	10.05 (0.45)	4×10^{-4} (0.04)	0.65	0.18
Tof	Alt.2	1.12 (0.24)	0.45 (0.19)	6.45 (0.14)	8.01 (0.18)	0.14 (0.03)	<0.001	0.01	0.59 (0.49)	1.84 (1.06)	6.62 (0.99)	9.06 (0.41)	0.07 (0.05)	0.07	0.03
Fmor	Alt.2	5×10^{-4} (2×10^{-4})	4×10^{-4} (2×10^{-4})	0.01 (3×10^{-4})	0.01 (3×10^{-4})	0.04 (0.02)	<0.001	0.03	-	0.02 (0.01)	0.07 (0.01)	0.09 (4×10^{-3})	-	-	0.22
Wean	Alt.2	0.1 (0.03)	4×10^{-7} (9×10^{-9})	2.72 (0.06)	2.82 (0.06)	0.04 (0.01)	<0.001	1	-	0.47 (0.7)	4.87 (0.72)	5.34 (0.23)	-	-	0.31
Gest	Alt.2	0.94 (0.13)	0.23 (0.08)	1.39 (0.03)	2.55 (0.08)	0.37 (0.04)	<0.001	2×10^{-3}	0.96 (0.34)	0.56 (0.66)	3.31 (0.6)	4.83 (0.23)	0.2 (0.07)	<0.001	0.3

¥Transformed log(trait + 1).† Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Heritability estimates are very low to moderate, if estimable, during the non-epidemic phase for the majority of reproductive performance traits, with values ranging from 0.02 (0.01) for *Mum* to 0.36 (0.04) for *Gest*. Only the *Gest* and *Tof* traits show more moderate heritability estimates in non-epidemic phase. With the exception of *Gest*, heritability estimates are higher in the epidemic phase than in the non-epidemic phase. The strongest increase was found for the *Wean* trait, which has an estimated heritability of 0.03 during the non-epidemic phase and a heritability of 0.41 during the epidemic phase according to the Basic model. These increases in heritability for disease traits in the epidemic phase occur as a result of a proportionally greater increase in additive genetic variance than in phenotypic variance. Phenotypic variance increases across all traits in the epidemic phase as compared to the non-epidemic phase.

When alternative model 1 (Alt.1) was applied to the epidemic phase data, (i.e. when the Epidemic ID is fitted as additional fixed effect to account for difference in the severity of the PRRSV outbreaks) heritability estimates decreased compared to those obtained by fitting the Basic Model. Heritability is only significant for *Mum, Still Alive* and *Gest* at $p < 0.05$. The environmental variance remains roughly the same between the basic and alternative 1 models, the reduction in heritability occurs as a result of a reduction in the additive genetic variance.

When alternative model 2 (Alt.2) was applied to the epidemic phase data, (i.e. when the epidemic specific rolling average trait mean is accounted for as a covariate in addition to epidemic ID), heritability estimates for reproductive traits decreased even further. Heritability became inestimable or not significantly different from zero for all traits except for *Gest*, for which heritability was approximately 0.2. In contrast, fitting the rolling trait average in the models of the non-epidemic phase data had hardly any noticeable effect on the heritability estimates.

G Matrix

The G matrix was calculated as per the method described in section 3.02.2, this is shown as a heat plot in Figure 3.3. Note the diagonal symmetry and maximum numbers ~1 lying on the diagonal. It shows, generally, higher relationships within genetic lines and lower levels between genetic lines.

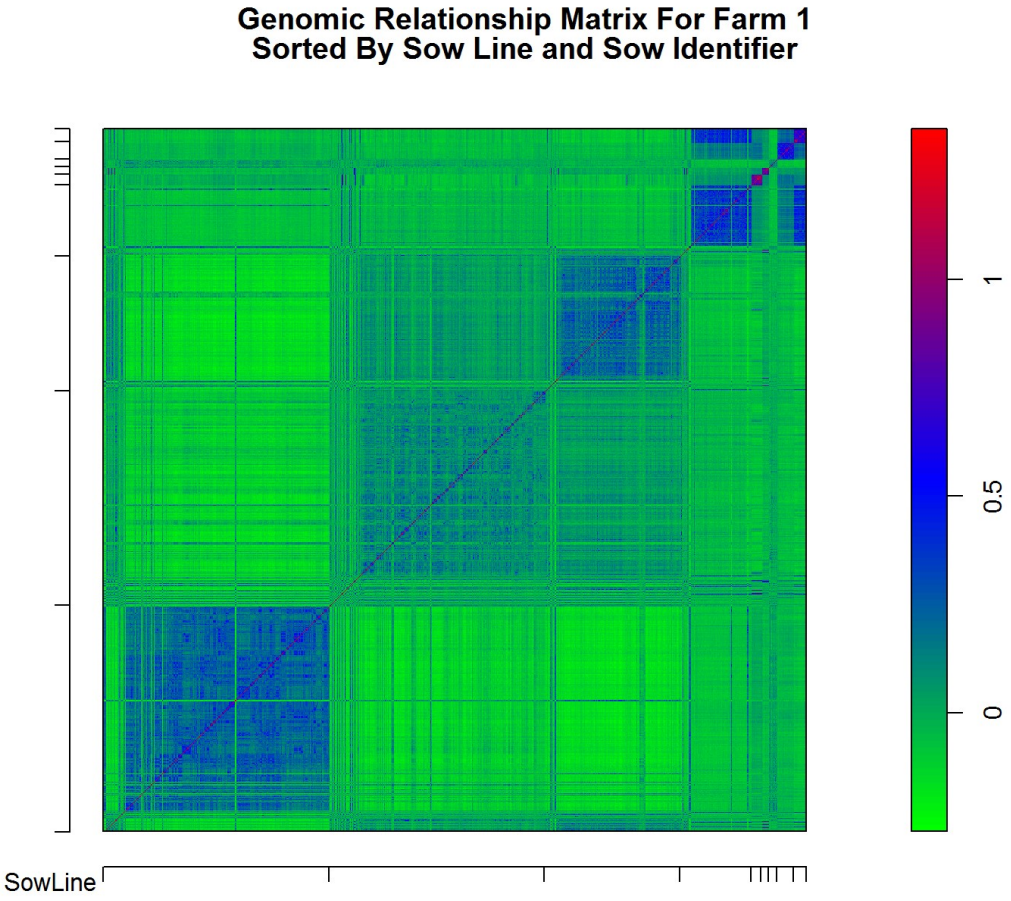

Figure 3.3 – Heat Plot of the Genomic Relationship Matrix for Farm 1
Heatmap of the genomic relationship matrix calculated using SNP data (see section 3.02.2). Ordered by sow line and sow identifier.

There are a number of inconsistencies within and between genetic lines. These were checked within the data and were found to be accurate based on the information provided. Erroneous values could however occur due to mislabelling of animals; or the observed patterns could be

an artefact of ordering within sow line, by sow identifier (which increase over time). This ordering by time means that changes in decision making along the course of breeding programmes could show as abrupt changes in the nature of relationships. Without further information to warrant the removal of these animals from the analysis all animals were retained and principal components explored for further consideration of population structure.

The differences in within line and between line relationships in Figure 3.3 suggests stratification in the population with distinct groups being identifiable. Whilst fitting sow line may partly capture this further consideration was made given the inconsistencies in sow line described above, and because of different degrees of relatedness between sow lines. The first 10 components were plotted to investigate those which show clustering by known sow line information, following this the first three principal components were selected for inclusion in the statistical model, as shown in Figure 3.4.

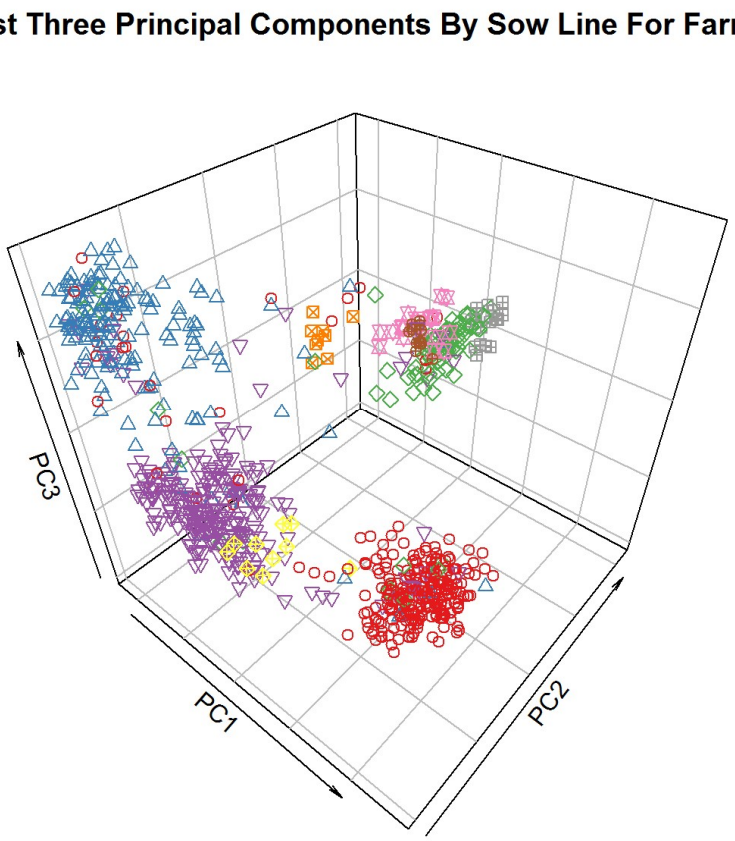

Figure 3.4 – First Three Principal Components by Sow Line for Farm 1

Sow lines are shown by distinct colours and shapes. Sow lines plotted using these principal components show clustering according to known sow line information, with some sow lines genetically more similar than others.

Variance Component and Heritability Estimates

For the repeated measures analysis, the additive genetic variance to standard error ratios were (generally) higher than for the single record analysis, with the repeated measures analysis showing lower heritability estimate standard errors. Marginal differences were seen in actual heritability estimate (<0.05). As such results are presented for the repeated measures analysis, summary statistics for these data are shown in Table 2.4.c.

Table 3.3 shows the variance component estimates, heritability and significance of the additive genetic variance of the reproductive traits under the repeated measures model for the non-epidemic phase and epidemic phase. Results from the single record analysis are shown in the appendix in Table A.6, summary statistics for these data are shown in Table 2.4.d.

Similar to the A matrix, estimation of a permanent environmental effect was conflicted by low numbers of repeated records for the epidemic phases. For estimates generated using the G matrix for non-epidemic phase data, significant ($p<-0.05$) permanent environmental effects could be estimated for the traits *Dead Alive Tof* and *Fmor* under the basic model. During the epidemic phase, significant permanent environmental effects could only be obtained for *Tof* and *Fmor* traits.

Table 3.3 – Farm 1 Variance Components and Heritability Estimated Using the G Matrix with a Repeated Measures Model for Non-Epidemic (Green) and Epidemic (Red) Phase

Trait	Model	Non-Epidemic							Epidemic						
		σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum¥	Basic	3×10^{-3} (2×10^{-3})	5×10^{-4} (2×10^{-3})	0.1 (3×10^{-3})	0.1 (3×10^{-3})	0.03 (0.02)	0.04	0.59	0.04 (0.04)	0.14 (0.1)	0.51 (0.1)	0.69 (0.04)	0.06 (0.06)	0.08	0.1
Still¥	Basic	2×10^{-3} (3×10^{-3})	4×10^{-3} (4×10^{-3})	0.18 (0.01)	0.19 (0.01)	0.01 (0.02)	0.31	0.14	0.07 (0.03)	2×10^{-7} (2×10^{-8})	0.38 (0.03)	0.45 (0.03)	0.16 (0.07)	<0.001	1
Dead¥	Basic	6×10^{-4} (4×10^{-3})	0.01 (0.01)	0.22 (0.01)	0.24 (0.01)	3×10^{-3} (0.02)	0.64	0.01	0.05 (0.04)	0.3 (0.09)	0.37 (0.08)	0.72 (0.04)	0.07 (0.06)	0.07	<0.001
Alive	Basic	0.01 (0.02)	0.06 (0.03)	1.2 (0.04)	1.26 (0.04)	0.01 (0.02)	0.51	0.02	1.4 (0.92)	6.4 (1.65)	6 (1.36)	13.8 (0.84)	0.1 (0.06)	0.02	<0.001
Tof	Basic	0.56 (0.25)	0.54 (0.23)	6.44 (0.23)	7.55 (0.25)	0.07 (0.03)	<0.001	0.01	0.91 (0.58)	2.69 (1.14)	5.11 (1.04)	8.71 (0.53)	0.1 (0.06)	0.02	0.01
Fmor	Basic	-	6×10^{-4} (2×10^{-4})	0.01 (4×10^{-4})	0.01 (4×10^{-4})	-	-	0.01	0.01 (0.01)	0.06 (0.01)	0.04 (0.01)	0.12 (0.01)	0.11 (0.06)	0.01	<0.001
Wean	Basic	0.04 (0.03)	2×10^{-7} (5×10^{-9})	2.15 (0.07)	2.19 (0.07)	0.02 (0.01)	0.07	1	2.97 (0.85)	1.31 (1.16)	5.34 (1.05)	9.62 (0.63)	0.31 (0.08)	<0.001	0.14
Gest	Basic	0.56 (0.12)	0.42 (0.09)	1.42 (0.05)	2.4 (0.1)	0.23 (0.04)	<0.001	<0.001	1.36 (0.46)	2×10^{-6} (1×10^{-7})	4.36 (0.38)	5.72 (0.36)	0.24 (0.07)	<0.001	1
Mum¥	Alt.1†								0.03 (0.04)	0.04 (0.11)	0.58 (0.11)	0.66 (0.04)	0.05 (0.05)	0.13	0.45
Still¥	Alt.1†								0.04 (0.03)	4×10^{-7} (3×10^{-8})	0.39 (0.03)	0.43 (0.03)	0.08 (0.06)	0.05	1
Dead¥	Alt.1†								0.02 (0.04)	0.17 (0.1)	0.48 (0.1)	0.67 (0.04)	0.03 (0.05)	0.34	0.07
Alive	Alt.1†								0.71 (0.79)	4.82 (1.94)	7.52 (1.76)	13.05 (0.78)	0.05 (0.06)	0.14	0.03
Tof	Alt.1†								0.63 (0.53)	2.74 (1.16)	5.16 (1.06)	8.52 (0.51)	0.07 (0.06)	0.06	0.01
Fmor	Alt.1†								0.01 (0.01)	0.05 (0.01)	0.06 (0.01)	0.11 (0.01)	0.07 (0.06)	0.09	0.01
Wean	Alt.1†								0.85 (0.56)	8×10^{-6} (6×10^{-7})	7.21 (0.58)	8.06 (0.48)	0.1 (0.07)	0.05	1
Gest	Alt.1†								1.27 (0.47)	0.38 (0.84)	3.98 (0.81)	5.63 (0.36)	0.23 (0.08)	<0.001	0.42

Trait	Model	Non-Epidemic							Epidemic						
		σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum¥	Alt.2	3×10^{-3} (2×10^{-3})	4×10^{-4} (2×10^{-3})	0.1 (3×10^{-3})	0.1 (3×10^{-3})	0.03 (0.02)	0.03	0.59	0.03 (0.03)	0.03 (0.1)	0.51 (0.1)	0.57 (0.03)	0.05 (0.05)	0.12	0.56
Still¥	Alt.2	1×10^{-3} (3×10^{-3})	4×10^{-3} (4×10^{-3})	0.18 (0.01)	0.18 (0.01)	0.01 (0.02)	0.43	0.15	-	0.05 (0.06)	0.3 (0.06)	0.35 (0.02)	-	-	0.36
Dead¥	Alt.2	1×10^{-3} (4×10^{-3})	0.01 (0.01)	0.22 (0.01)	0.23 (0.01)	4×10^{-3} (0.02)	0.52	0.01	0.01 (0.03)	0.17 (0.08)	0.36 (0.08)	0.54 (0.03)	0.01 (0.05)	0.53	0.06
Alive	Alt.2	0.01 (0.02)	0.05 (0.03)	1.18 (0.04)	1.25 (0.04)	0.01 (0.02)	0.32	0.04	0.18 (0.57)	3.97 (1.68)	6.45 (1.55)	10.6 (0.63)	0.02 (0.05)	0.46	0.06
Tof	Alt.2	0.61 (0.26)	0.51 (0.23)	6.44 (0.23)	7.56 (0.25)	0.08 (0.03)	<0.001	0.01	0.61 (0.52)	2.49 (1.16)	5.23 (1.08)	8.32 (0.5)	0.07 (0.06)	0.06	0.02
Fmor	Alt.2	4×10^{-9} (2×10^{-10})	5×10^{-4} (2×10^{-4})	0.01 (4×10^{-4})	0.01 (3×10^{-4})	0e+00 (0e+00)	1	0.01	2×10^{-3} (5×10^{-3})	0.05 (0.01)	0.04 (0.01)	0.09 (0.01)	0.03 (0.05)	0.3	<0.001
Wean	Alt.2	0.04 (0.03)	2×10^{-7} (6×10^{-9})	2.14 (0.07)	2.18 (0.07)	0.02 (0.01)	0.05	1	0.27 (0.35)	1.05 (0.83)	3.86 (0.78)	5.18 (0.3)	0.05 (0.07)	0.25	0.13
Gest	Alt.2	0.57 (0.12)	0.43 (0.09)	1.4 (0.05)	2.4 (0.1)	0.24 (0.04)	<0.001	<0.001	1.29 (0.46)	0.49 (0.82)	3.73 (0.79)	5.51 (0.35)	0.23 (0.08)	<0.001	0.36

¥Transformed log(trait + 1).† Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Under the basic model for the non-epidemic phase, no significant heritable component could be estimated for the majority of traits except for *Mum*, *Tof* and *Gest*. In contrast, in epidemic phase basic model heritability estimates are significantly different from zero ($p<0.05$) for all traits except *Mum* and *Dead* which only indicate a significant heritable component at $p<0.1$. A moderate heritability estimate is generated for the *Wean* trait at $h^2=0.31$. Under the G matrix, in epidemic phase h^2 estimates are comparable to the A matrix estimates for *Still* at 0.16 (*cf*. 0.18 under the A-matrix) and Gest at 0.24 (*cf*. 0.21 under the A-matrix). In the other traits indicting a significant heritable effect in the epidemic phase under the G matrix (*Mum*, *Alive*, *Tof* and *Fmor*), a slightly smaller through not significant difference in h^2 estimate were obtained with the G Matrix.

Compared to the Basic Model, alternative model 1 applied to the epidemic phase data showed that heritability is considerably reduced for most traits as a result of a reduction in the genetic variance component. Heritability is estimable and significant only for the traits *Still, Wean* and *Gest.* The *Still, and Wean* disease indicator traits still have a higher heritability estimate in the epidemic phase compared to the basic model non-epidemic phase. In line with the results using the A matrix, heritability estimates for most reproductive traits decreased further when the alternative model 2 (Alt.2) was fitted to the epidemic phase data. Also in line with the pedigree-based models, fitting the rolling trait average in the models of the non-epidemic phase data had hardly any noticeable effect on the heritability estimates.

Across all the traits a reduction is seen in the actual additive genetic variance estimated using the G matrix as compared to the A matrix

Joint Estimation of the Pedigree and Genomic Variance Components

To investigate whether the variance explained by the two separate matrices (A and G) showed evidence of confounding the two matrices were fit simultaneously to compare their relative variance estimates. Heritability was calculated as the sum of the additive genetic

variance components calculated under the A and G matrices divided, by the phenotypic variance. Results are presented using a repeated measures model on multiple record data in Table 3.4.

Table 3.4 – Farm 1 Variance Components and Heritability Estimated Jointly for the A & G Matrices using a Repeated Measures Model with Non-Epidemic Phase (Green) and Epidemic Phase (Red) Data

Trait	Model	Non-Epidemic Phase					Epidemic Phase				
		σ^2_A (SE)		LRT P Value		h^2 (SE)	σ^2_A (SE)		LRT P Value		h^2 (SE)
		G Matrix	A Matrix	G Matrix	A Matrix		G Matrix	A Matrix	G Matrix	A Matrix	
Mum¥	Basic	1×10^{-3} (2×10^{-3})	3×10^{-3} (2×10^{-3})	0.31	0.02	0.04 (0.02)	0.01 (0.04)	0.08 (0.07)	0.54	0.07	0.13 (0.08)
Still¥	Basic	2×10^{-8} (6×10^{-10})	5×10^{-3} (4×10^{-3})	1	0.07	0.03 (0.02)	0.02 (0.04)	0.11 (0.06)	0.44	0.03	0.27 (0.1)
Dead¥	Basic	3×10^{-8} (1×10^{-9})	0.01 (0.01)	1	0.02	0.03 (0.02)	7×10^{-8} (8×10^{-9})	0.14 (0.07)	1	0.02	0.2 (0.1)
Alive	Basic	7×10^{-7} (3×10^{-8})	0.02 (0.03)	1	0.16	0.02 (0.02)	0.11 (0.98)	3.12 (1.69)	1	0.01	0.23 (0.1)
Tof	Basic	0.42 (0.26)	0.58 (0.33)	0.02	0.01	0.13 (0.04)	0.36 (0.65)	1.39 (1.02)	0.31	0.06	0.2 (0.1)
Fmor	Basic	1×10^{-9} (4×10^{-11})	2×10^{-4} (2×10^{-4})	1	0.17	0.02 (0.02)	2×10^{-3} (0.01)	0.03 (0.01)	0.58	0.02	0.24 (0.1)
Wean	Basic	0.33 (0.12)	0.59 (0.17)	<0.001	<0.001	0.38 (0.06)	0.9 (0.56)	1.37 (0.7)	0.04	0.01	0.38 (0.11)
Gest	Basic	1×10^{-3} (2×10^{-3})	3×10^{-3} (2×10^{-3})	0.31	0.02	0.04 (0.02)	3×10^{-3} (0.04)	0.07 (0.06)	0.74	0.1	0.1 (0.08)
Mum¥	Alt.1†						2×10^{-7} (2×10^{-8})	0.09 (0.04)	1	0.03	0.21 (0.1)
Still¥	Alt.1†						1×10^{-7} (2×10^{-8})	0.09 (0.06)	1	0.05	0.13 (0.09)
Dead¥	Alt.1†						7×10^{-6} (8×10^{-7})	2.13 (1.36)	1	0.03	0.16 (0.1)
Alive	Alt.1†						0.43 (0.65)	0.86 (0.96)	0.24	0.17	0.15 (0.1)
Tof	Alt.1†						6×10^{-8} (7×10^{-9})	0.02 (0.01)	1	0.03	0.17 (0.1)
Fmor	Alt.1†						0.83 (0.55)	1.11 (0.68)	0.05	0.03	0.34 (0.11)
Wean	Alt.1†						9×10^{-8} (8×10^{-9})	0.04 (0.04)	1	0.17	0.07 (0.07)
Gest	Alt.1†						2×10^{-8} (2×10^{-9})	5×10^{-3} (0.03)	1	0.62	0.01 (0.07)
Mum¥	Alt.2	6×10^{-8} (2×10^{-9})	0.01 (0.01)	1	0.04	0.03 (0.02)	9×10^{-8} (8×10^{-9})	0.02 (0.04)	1	0.29	0.04 (0.07)
Still¥	Alt.2	0.01 (0.03)	0.01 (0.03)	0.48	0.29	0.02 (0.02)	3×10^{-6} (3×10^{-7})	0.57 (0.87)	1	0.23	0.05 (0.08)
Dead¥	Alt.2	0.46 (0.26)	0.58 (0.33)	0.02	0.01	0.14 (0.04)	0.28 (0.62)	1.08 (0.99)	0.36	0.12	0.16 (0.1)
Alive	Alt.2	3×10^{-9} (9×10^{-11})	1×10^{-4} (2×10^{-4})	1	0.23	0.01 (0.02)	3×10^{-8} (3×10^{-9})	0.01 (0.01)	1	0.18	0.07 (0.08)
Tof	Alt.2	0.33 (0.12)	0.59 (0.17)	<0.001	<0.001	0.39 (0.06)	0.78 (0.53)	1.12 (0.67)	0.06	0.02	0.33 (0.11)
Fmor	Alt.2	1×10^{-3} (2×10^{-3})	3×10^{-3} (2×10^{-3})	0.31	0.02	0.04 (0.02)	0.01 (0.04)	0.08 (0.07)	0.54	0.07	0.13 (0.08)
Wean	Alt.2	2×10^{-8} (6×10^{-10})	5×10^{-3} (4×10^{-3})	1	0.07	0.03 (0.02)	0.02 (0.04)	0.11 (0.06)	0.44	0.03	0.27 (0.1)
Gest	Alt.2	3×10^{-8} (1×10^{-9})	0.01 (0.01)	1	0.02	0.03 (0.02)	7×10^{-8} (8×10^{-9})	0.14 (0.07)	1	0.02	0.2 (0.1)

¥Transformed log(trait + 1).† Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Generally additive genetic variance is primarily seen under either the A or G matrix. This indicates that the two matrices are not independent in terms of representing the relationships in the population, and that most of the genetic variance in the trait of interest is captured by the A matrix. This demonstrates, as expected, considerable confounding of the relationships described by the two matrices. Whilst some inflation of the heritability is observed (as compared to the A matrix estimate) this is within the standard error suggesting no significant change in overall estimate.

3.03.2 Farm 2

There was no pedigree information available for farm 2. Considerably fewer numbers of records are available than for the analysis in chapter 2 for the generation of the G matrix; requiring animals that are both genotyped and pass genetic QC. Multiple records were only available for non-epidemic phase data and epidemic including unknown phase data, for which single record datasets were also created. In keeping with the aims of the research, in the single record data litter records were favoured from the PRRSV confirmed epidemic leading to an unequal number of litters from the PRRSV confirmed epidemic (n=276) as compared to the unknown epidemic (n=53). This reduction in records in the single record datasets had a slight impact on the trait means, as shown in Table 3.5.

Table 3.5 – Summary Statistics Showing Counts and Trait Means for Farm 2 Datasets

Level	Records	Sows	*Mum*	*Still*	*Dead*	*Alive*	*Tof*	*Fmor*	*Wean*	*Gest*
a. Genotyped† Animals Single Records										
Non-epidemic Phase	585	585	0.27 (0.02)	0.76 (0.05)	1.03 (0.05)	12.58 (0.12)	13.6 (0.12)	0.07 (4×10^{-3})	115.12 (0.05)	585
Epidemics Total	329	329	0.6 (0.08)	1.3 (0.1)	1.9 (0.13)	12.34 (0.18)	14.24 (0.17)	0.13 (9×10^{-3})	115.7 (0.09)	329
Epidemic 4 (confirmed)	276	276	0.6 (0.09)	0.92 (0.19)	1.53 (0.27)	11.96 (0.43)	13.49 (0.37)	0.11 (0.02)	114.91 (0.19)	276
Epidemic 5 (unconfirmed)	53	53	0.6 (0.2)	1.37 (0.11)	1.97 (0.14)	12.42 (0.2)	14.39 (0.19)	0.13 (1×10^{-2})	115.85 (0.1)	53
b. Genotyped† Animal Multiple Records										
Non-epidemic Phase	1684	585	0.26 (0.01)	0.77 (0.03)	1.04 (0.03)	12.61 (0.07)	13.65 (0.07)	0.07 (2×10^{-3})	115.57 (0.03)	1684
Epidemics Total	380	329	0.58 (0.07)	1.31 (0.09)	1.89 (0.12)	12.27 (0.17)	14.17 (0.16)	0.13 (8×10^{-3})	115.57 (0.08)	380
Epidemic 4 (confirmed)	276	276	0.6 (0.09)	1.15 (0.16)	1.69 (0.21)	11.89 (0.33)	13.59 (0.3)	0.12 (0.02)	114.84 (0.14)	276
Epidemic 5 (unconfirmed)	104	104	0.54 (0.13)	1.37 (0.11)	1.97 (0.14)	12.42 (0.2)	14.39 (0.19)	0.13 (1×10^{-2})	115.85 (0.1)	104

†Genotyped and passing genetic quality control. Summary statistics for data under different partitions showing number of records, number of sows and means with standard error of the mean for the reproductive performance traits considered in this study. Analyses excluding the unknown epidemic use non-epidemic phase and epidemic 4 only.

G-Matrix

A plot of the genomic relationship matrix for farm 2 animals is shown in Figure 3.6. Note a lack of between line distinction. This is as animals belong to the same crosses from the same breeding stock with only slight differences in breeding objectives.

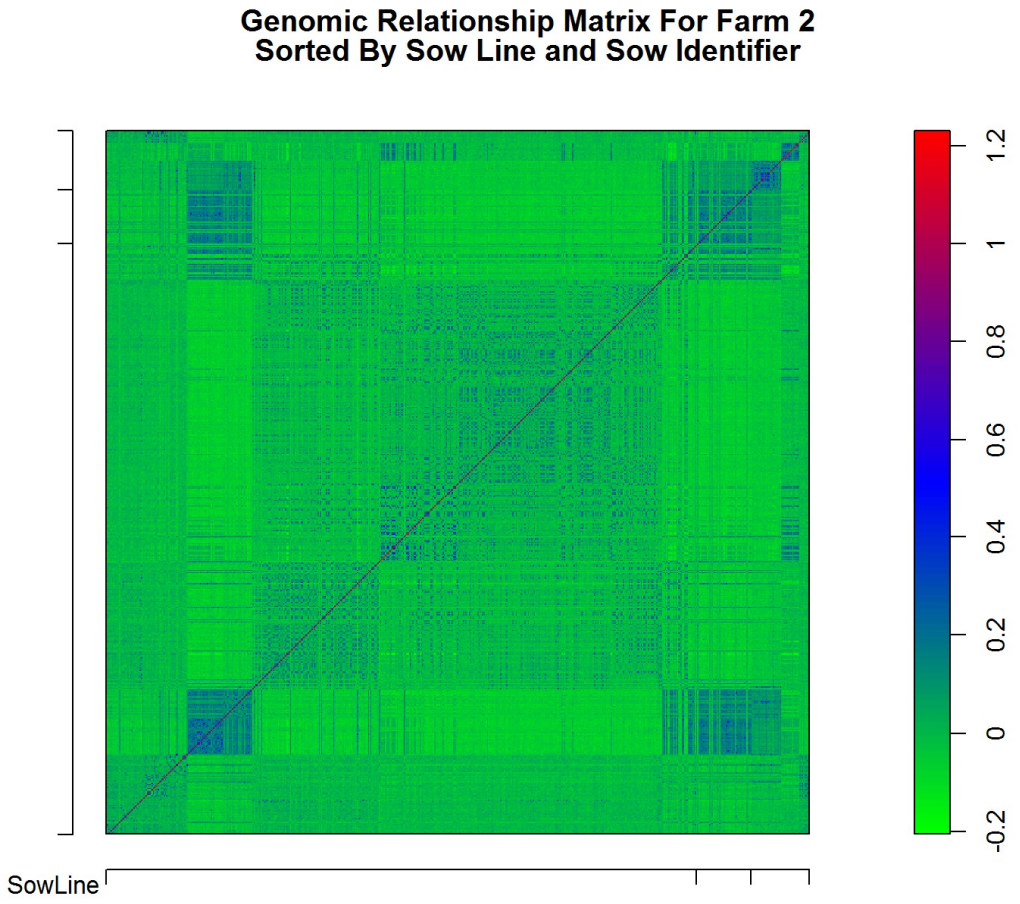

Figure 3.5 – Heatmap of the Genomic Relationship Matrix for Farm 2
Heatmap of the genomic relationship matrix calculated using SNP data (see section 3.02.2). Ordered by sow line and sow identifier.

The dominating patterns observed in the data are an artefact of sorting by sow identifier (which increase over time). This ordering by time means that changes in decision making along the time-course of the breeding programmes, show as distinct changes to relationships within lines. The comparatively minor distinctions between lines become more apparent if

we zoom in on a smaller subset of animals shown in Figure 3.6, intended to equalise the representation of each line, reducing the number of animals in the plot and allowing better contrasts to be made by individual. Here only the first 50 animals from the first sow line are included in the plot.

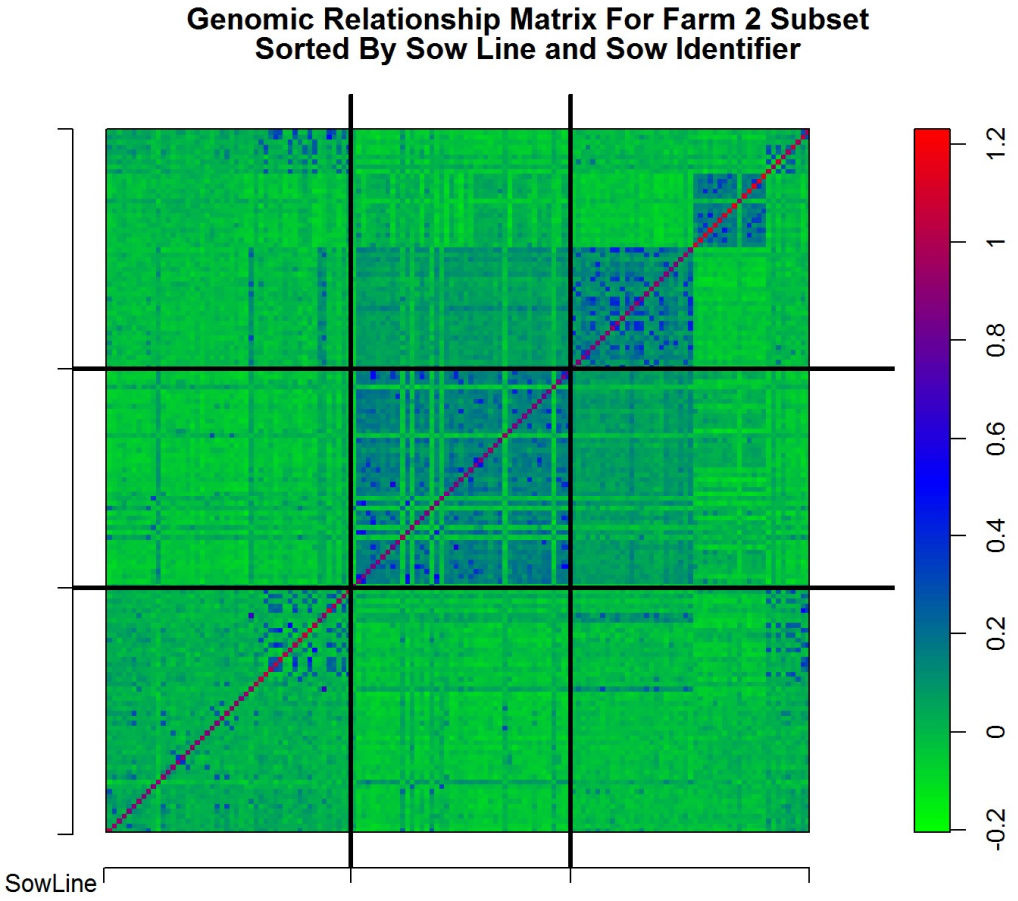

Figure 3.6 – Heatmap of the Genomic Relationship Matrix for Subset of Farm 2 Animals
Heatmap of the genomic relationship matrix calculated using SNP data (see section 3.02.2). Shown ordered by sow line and sow identifier. First 50 animals from the first line are included against all animals from second and third lines.

Minor differences can now be seen in IBD between the lines, while the overriding patterns dominating the matrix are still seen as families.

A plot of the first three principal components for farm 2 is shown in Figure 3.7. These partition a cluster which appears to be a within line family rather than a between line cluster.

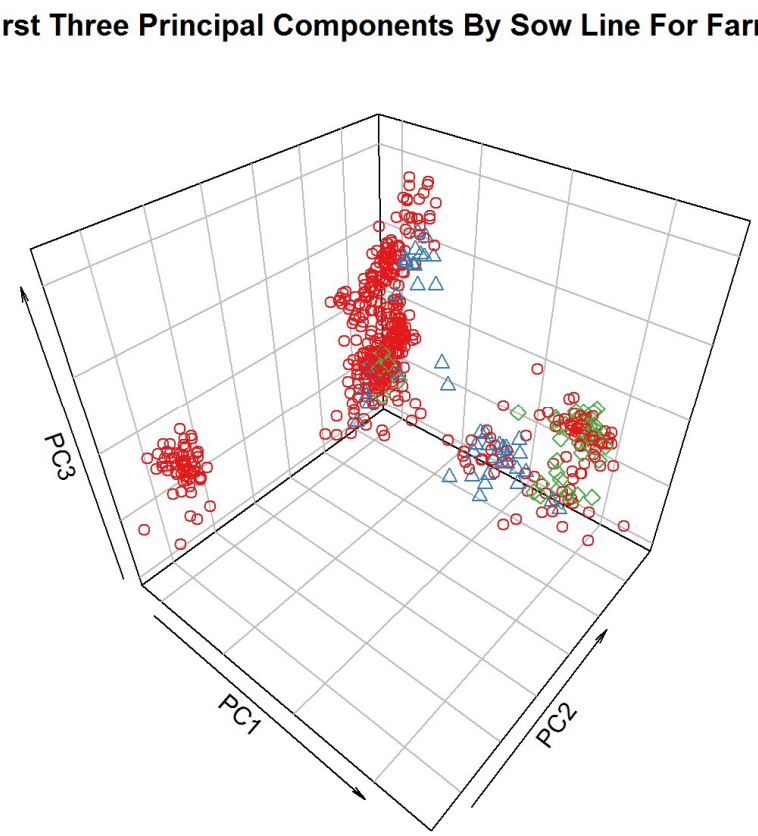

Figure 3.7 – First Three Principal Components by Sow Line for Farm 2
Sow lines are shown by distinct colours and shapes.

This is consistent with the heat plot of the kindship matrix in Figure 3.6 which showed minor distinction between breeding line groups. The patterns dominating the IBS matrix plots are "family". It is therefore expected that principal components from this data would isolate family differences as opposed to line differences. As such principal components were not used in estimating variance components for Farm 2.

Variance Component and Heritability Estimates

Variance component estimates, heritability and LRT p-values for random effects are shown in Table 3.6, for the epidemic and non-epidemic phases. No repeated records per sow were

available in the epidemic phase (Epidemic 4 in Table 3.5) as such no permanent environmental effect could be fitted in the model. In the non-epidemic phase, a considerable reduction is seen in the number of records between the multiple record data (n=1,684) as compared to the single record data (n=585) as such the repeated measures model is presented. Alternative model 1 where the individual epidemic effect is considered was not applicable to this analysis as only one epidemic is fitted.

Table 3.6 – Farm 2 Variance Components and Heritability Estimated Using a G Matrix with a Repeated Measures Model for Non-Epidemic Phase (Green) and Single Record per Sow in Epidemic Phase (Red)

Trait	Model	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_A (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum¥	Basic	4×10^{-3} (3×10^{-3})	2×10^{-3} (3×10^{-3})	0.1 (4×10^{-3})	0.11 (4×10^{-3})	0.04 (0.03)	0.03	0.37	0.03 (0.04)	0.22 (0.03)	0.25 (0.02)	0.13 (0.14)	0.16
Still¥	Basic	0.01 (0.01)	4×10^{-8} (1×10^{-9})	0.21 (0.01)	0.22 (0.01)	0.06 (0.02)	2×10^{-4}	1	0.06 (0.05)	0.29 (0.05)	0.35 (0.03)	0.18 (0.14)	0.06
Dead¥	Basic	0.02 (0.01)	1×10^{-7} (4×10^{-9})	0.23 (0.01)	0.25 (0.01)	0.06 (0.02)	2×10^{-5}	1	-	0.42 (0.04)	0.42 (0.04)	-	-
Alive	Basic	0.09 (0.03)	5×10^{-8} (2×10^{-9})	1.33 (0.05)	1.42 (0.05)	0.06 (0.02)	1×10^{-4}	1	0.58 (0.63)	4.18 (0.63)	4.76 (0.43)	0.12 (0.13)	0.13
Tof	Basic	0.96 (0.4)	0.88 (0.36)	7.14 (0.3)	8.97 (0.34)	0.11 (0.04)	1×10^{-4}	0.01	0.22 (0.94)	8.78 (1.11)	9 (0.79)	0.02 (0.11)	0.54
Fmor	Basic	6×10^{-4} (2×10^{-4})	-	0.01 (3×10^{-4})	0.01 (3×10^{-4})	0.08 (0.03)	5×10^{-4}	-	2×10^{-3} (3×10^{-3})	0.02 (3×10^{-3})	0.02 (2×10^{-3})	0.1 (0.13)	0.2
Gest	Basic	0.4 (0.12)	0.54 (0.1)	0.81 (0.03)	1.75 (0.09)	0.23 (0.06)	1×10^{-5}	1×10^{-10}	0.36 (0.36)	1.85 (0.32)	2.21 (0.21)	0.16 (0.15)	0.19
Mum¥	Alt.2	4×10^{-3} (3×10^{-3})	2×10^{-3} (3×10^{-3})	0.1 (4×10^{-3})	0.11 (4×10^{-3})	0.04 (0.03)	0.02	0.33	0.04 (0.04)	0.2 (0.03)	0.25 (0.02)	0.17 (0.15)	0.11
Still¥	Alt.2	0.02 (0.01)	4×10^{-8} (2×10^{-9})	0.2 (0.01)	0.22 (0.01)	0.07 (0.02)	4×10^{-5}	1	0.07 (0.05)	0.29 (0.05)	0.35 (0.03)	0.19 (0.14)	0.06
Dead¥	Alt.2	0.02 (0.01)	1×10^{-7} (5×10^{-9})	0.23 (0.01)	0.25 (0.01)	0.07 (0.02)	4×10^{-6}	1	0.01 (0.05)	0.41 (0.05)	0.42 (0.04)	0.02 (0.11)	0.59
Alive	Alt.2	0.1 (0.04)	6×10^{-8} (2×10^{-9})	1.3 (0.05)	1.4 (0.05)	0.07 (0.02)	3×10^{-5}	1	0.8 (0.7)	3.96 (0.64)	4.76 (0.44)	0.17 (0.14)	0.07
Tof	Alt.2	0.94 (0.4)	0.89 (0.36)	7.15 (0.3)	8.97 (0.34)	0.1 (0.04)	2×10^{-4}	0.01	0.25 (0.97)	8.78 (1.13)	9.03 (0.8)	0.03 (0.11)	0.51
Fmor	Alt.2	8×10^{-4} (2×10^{-4})	9×10^{-10} (3×10^{-11})	0.01 (3×10^{-4})	0.01 (3×10^{-4})	0.1 (0.03)	4×10^{-6}	1	4×10^{-3} (4×10^{-3})	0.02 (3×10^{-3})	0.02 (2×10^{-3})	0.17 (0.14)	0.08
Gest	Alt.2	0.38 (0.13)	0.57 (0.1)	0.79 (0.03)	1.74 (0.09)	0.22 (0.07)	7×10^{-5}	3×10^{-11}	0.33 (0.35)	1.88 (0.32)	2.21 (0.21)	0.15 (0.15)	0.21

¥Transformed log(trait + 1).† Alt.1 model not applicable for farm 2 data given the existence of only one epidemic. - denotes inestimable.

In the non-epidemic phase, all traits show a significant heritable effect according to the LRT ($p<0.05$), according to both, the Basic and Alt2 model (Table 9). For the epidemic phase, heritability estimates were not significantly different from zero for all reproductive traits considered, though indicated for the *Still* trait at P=0.06. Actual heritability estimates, as well as standard errors, were larger for the epidemic phase, suggesting insufficient statistical power to detect heritability of reproductive traits in farm 2.

Interestingly where a significant additive genetic component is indicated ($P<0.1$) for the *Still* trait and using the Alt.2 model for *Alive* and *Fmor* traits, a proportionally greater increase is seen using the Alt.2 model as compared to the basic model. This is contrary to the pattern observed in the farm 1 estimates where a heritable effect could not be demonstrated under the Alt.2 model.

Results from the single record analysis of the non-epidemic phase is shown in the appendix Table A.7. As expected calculation of an additive genetic variance component is limited by the considerable reduction in the number of records, with an increase observed in the standard error of the estimate. Only the *Gest* trait shows a significant ($P<0.05$) heritable component above zero in these data (observed in both Basic and Alt.2 models applicable to these data). A heritable effect is indicated in the *Still* and *Dead* traits at $P<0.1$.

Variance components and heritability was also estimated in data including the unknown epidemic. Given the low numbers of repeated records a significant permanent environmental effect could not be demonstrated in these data. Results from the Single record analysis are shown in Table 3.7. Summary statistics for this dataset are shown in Table 3.5.a.

Table 3.7 – Farm 2 Variance Components and Heritability Estimated Using the G Matrix Single Record Data for Epidemic Including Unknown Phase

Trait	Model	σ^2_A (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	σ2A LRT P
Mum¥	Basic	0.03 (0.03)	0.23 (0.03)	0.25 (0.02)	0.1 (0.11)	0.18
Still¥	Basic	0.06 (0.04)	0.28 (0.04)	0.34 (0.03)	0.18 (0.12)	0.03
Dead¥	Basic	0.01 (0.04)	0.4 (0.05)	0.42 (0.03)	0.03 (0.1)	0.47
Alive	Basic	0.71 (0.57)	3.95 (0.54)	4.66 (0.39)	0.15 (0.12)	0.04
Tof	Basic	0.79 (1)	8.01 (1.03)	8.81 (0.72)	0.09 (0.11)	0.17
Fmor	Basic	3×10^{-3} (3×10^{-3})	0.02 (3×10^{-3})	0.02 (2×10^{-3})	0.13 (0.12)	0.08
Gest	Basic	0.09 (0.26)	2.06 (0.27)	2.16 (0.18)	0.04 (0.12)	0.5
Mum¥	Alt.1	0.02 (0.03)	0.23 (0.03)	0.25 (0.02)	0.08 (0.11)	0.25
Still¥	Alt.1	0.06 (0.05)	0.28 (0.04)	0.34 (0.03)	0.18 (0.12)	0.03
Dead¥	Alt.1	0.02 (0.04)	0.4 (0.05)	0.42 (0.03)	0.04 (0.1)	0.43
Alive	Alt.1	0.74 (0.58)	3.92 (0.55)	4.66 (0.39)	0.16 (0.12)	0.04
Tof	Alt.1	0.75 (1.01)	8.08 (1.04)	8.82 (0.72)	0.08 (0.11)	0.19
Fmor	Alt.1	3×10^{-3} (3×10^{-3})	0.02 (3×10^{-3})	0.02 (2×10^{-3})	0.14 (0.12)	0.07
Gest	Alt.1	0.15 (0.27)	2.01 (0.27)	2.16 (0.18)	0.07 (0.12)	0.37
Mum¥	Alt.2	0.03 (0.03)	0.21 (0.03)	0.24 (0.02)	0.13 (0.13)	0.16
Still¥	Alt.2	0.06 (0.04)	0.28 (0.04)	0.34 (0.03)	0.17 (0.13)	0.05
Dead¥	Alt.2	0.02 (0.04)	0.4 (0.05)	0.41 (0.03)	0.04 (0.1)	0.41
Alive	Alt.2	0.82 (0.6)	3.82 (0.55)	4.64 (0.39)	0.18 (0.12)	0.03
Tof	Alt.2	0.73 (1.01)	8.13 (1.05)	8.86 (0.73)	0.08 (0.11)	0.2
Fmor	Alt.2	4×10^{-3} (3×10^{-3})	0.02 (3×10^{-3})	0.02 (2×10^{-3})	0.16 (0.12)	0.06
Gest	Alt.2	0.06 (0.24)	2.03 (0.26)	2.08 (0.17)	0.03 (0.12)	0.64

¥Transformed log(trait + 1). - denotes inestimable.

An improvement is seen when the unknown epidemic is included, with more disease traits showing a significant genetic effect. Heritability estimates significantly different from zero ($p<0.05$) are found for the *Still* and *Alive* traits and indicated for *Fmor* at $p=0.08$ with the basic model. For these traits when the epidemic effect was taken into account (Alt.1) the heritability increased slightly. When a dynamic by epidemic trend is taken into (Alt.2) a heritable effect was also demonstrated for *Still, Alive* and *Fmor* traits ($p<0.06$), the heritability estimates increasing marginally for the *Alive* and *Fmor* traits.

For the traits *Still, Alive* and *Fmor* a common heritability is demonstrated facing reproductive challenge from (most likely) different etiological agents (Table 3.7).

Results are shown in the appendix Table A.8 for the repeated measures analysis for the epidemic including unknown phase. An additional 104 litter records from the unknown epidemic (5) are added to the 276 records from epidemic 4 shown in Table 3.5.b. In these data, a consistent heritability of 0.13 (s.e. 0.09) was demonstrated for the Still trait in the Basic ($p=0.04$), Alternative 1 ($p=0.06$) and Alternative 2 ($p=0.07$) models.

3.03.3 Joint Analysis

In the combined data analysis 3,903 records from 1,292 sows were available for non-epidemic phase analysis.

G-Matrix

The G matrix calculated for the joint data is shown in Figure 3.8. This shows the same patterns from the two separate matrices combined and identifies some of the sow lines in farm 1 were crossed to create the sow lines on farm 2.

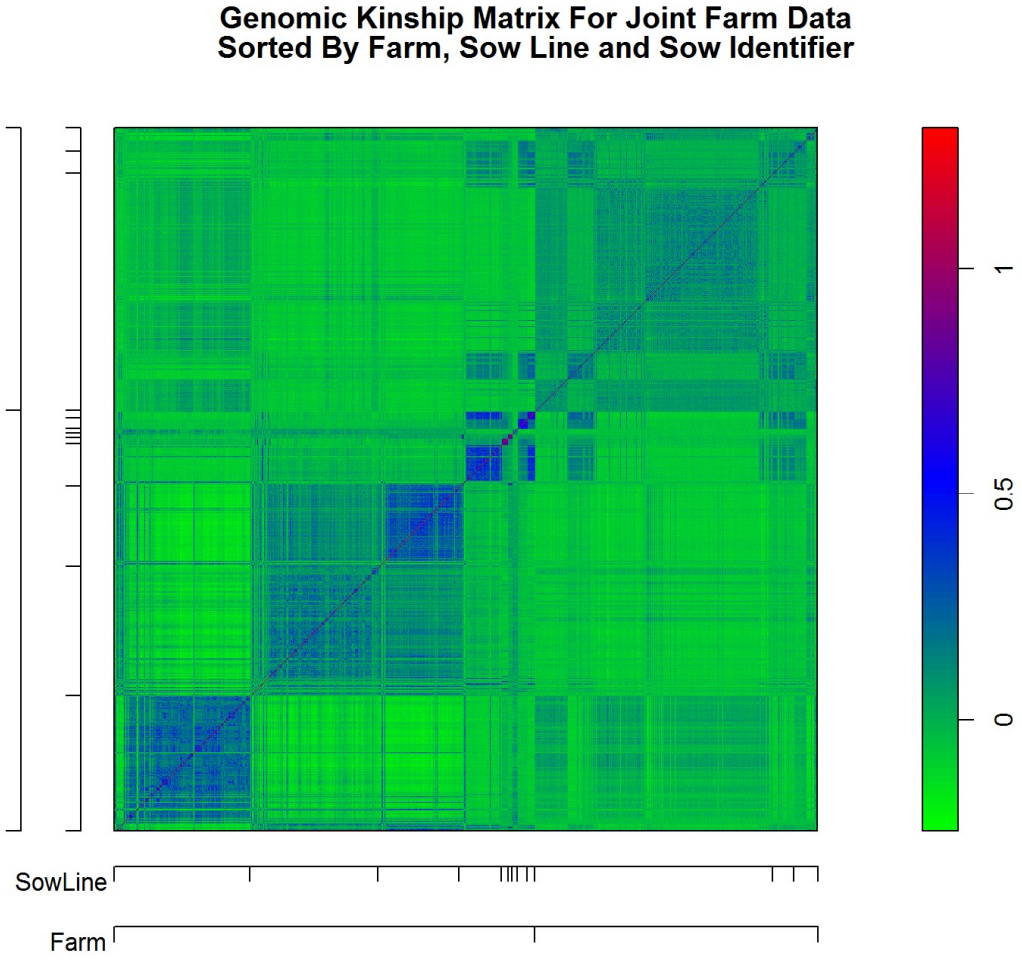

Figure 3.8 – Heatmap of the Genomic Relationship Matrix for Joint Farm Data

Heatmap of the genomic relationship matrix calculated using SNP data (see section 3.02.2). Ordered by sow line and sow identifier.

Using the same methodology as applied to farm 1 the first three principal components were fitted as additional covariates in the model. These are shown plotted in Figure 3.9.

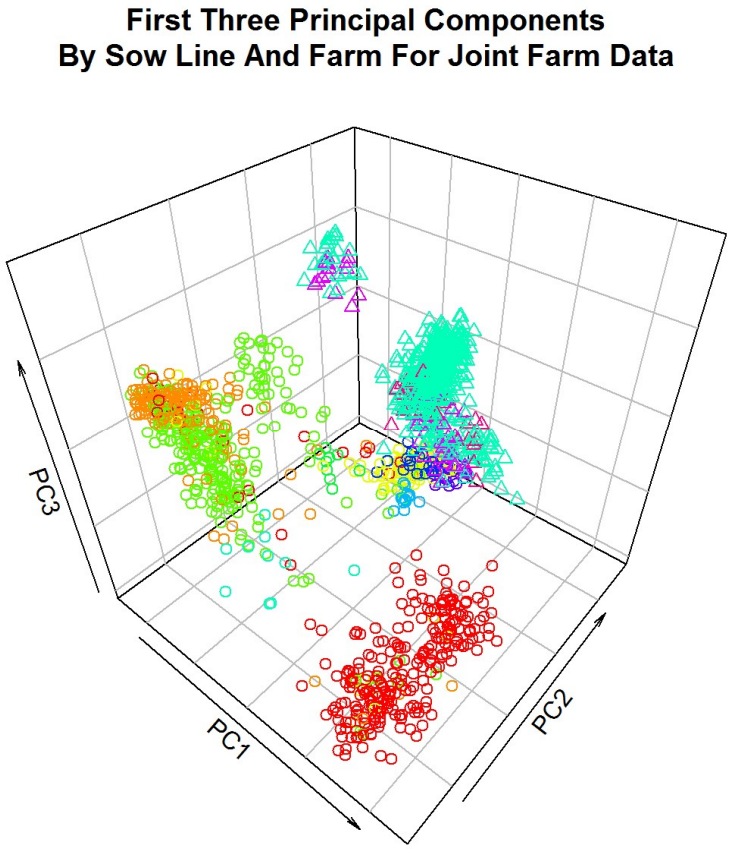

Figure 3.9 – First Three Principal Components by Sow Line for Joint Farm Data
Farm 1 (○) and Farm 2 (△) principle components on SNP data different, sow lines are shown with different colours.

Clustering can be seen between farms though some overlap is observed among sow lines of different farms, consistent with sow lines on farm 2 belonging to the same cross as some animals on farm 1.

Variance Components and Heritability Estimates

While similar problems, concerning repeated measures, are observed in the joint farm analysis as are seen in the individual farm analyses, generally larger additive genetic variance to standard error ratios are observed for epidemic phase estimated under the repeated measures analysis as compared to the single record analysis. As before, variance

components, heritability and random effect LRT p-values are shown in Table 3.8 for both non-epidemic and epidemic phases under the repeated measures model, as these indicate increased power.

Results from the single record analysis are shown in the appendix Table A.9.

Table 3.8 – Joint Farm Variance Components and Heritability Estimated Using the G Matrix with a Repeated Measures Model for Non-Epidemic (Green) and Epidemic (Red) Phase

Trait	Model	σ^2_G	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_G	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum	Basic	4×10^{-3} (2×10^{-3})	4×10^{-4} (2×10^{-3})	0.1 (3×10^{-3})	0.1 (2×10^{-3})	0.04 (0.02)	8×10^{-4}	0.54	0.03 (0.03)	0.09 (0.08)	0.44 (0.08)	0.56 (0.03)	0.06 (0.05)	0.06	0.1
Still	Basic	4×10^{-3} (3×10^{-3})	0.01 (4×10^{-3})	0.19 (0.01)	0.2 (5×10^{-3})	0.02 (0.01)	0.08	0.08	0.08 (0.03)	2×10^{-7} (1×10^{-8})	0.35 (0.02)	0.43 (0.02)	0.2 (0.06)	1×10^{-4}	1
Dead	Basic	5×10^{-3} (4×10^{-3})	0.01 (4×10^{-3})	0.23 (0.01)	0.24 (0.01)	0.02 (0.02)	0.07	0.02	0.04 (0.03)	0.26 (0.08)	0.34 (0.07)	0.65 (0.03)	0.07 (0.05)	0.05	1×10^{-3}
Alive	Basic	0.02 (0.02)	0.05 (0.02)	1.26 (0.03)	1.33 (0.03)	0.02 (0.01)	0.1	0.03	0.81 (0.58)	4.4 (1.33)	5.68 (1.21)	10.89 (0.55)	0.07 (0.05)	0.04	2×10^{-3}
Tof	Basic	0.69 (0.21)	0.77 (0.2)	6.73 (0.18)	8.19 (0.21)	0.08 (0.03)	8×10^{-6}	1×10^{-5}	0.93 (0.51)	2.72 (1.13)	5.15 (1.05)	8.8 (0.45)	0.11 (0.06)	0.01	0.01
Fmor	Basic	9×10^{-5} (1×10^{-4})	4×10^{-4} (2×10^{-4})	0.01 (3×10^{-4})	0.01 (2×10^{-4})	0.01 (0.01)	0.26	0.01	0.01 (5×10^{-3})	0.04 (0.01)	0.04 (0.01)	0.09 (5×10^{-3})	0.1 (0.05)	0.01	9×10^{-5}
Gest	Basic	0.54 (0.09)	0.46 (0.06)	1.18 (0.03)	2.19 (0.07)	0.25 (0.04)	$<3\times10^{-16}$	$<3\times10^{-16}$	0.94 (0.31)	1×10^{-6} (8×10^{-8})	3.72 (0.27)	4.66 (0.24)	0.2 (0.06)	8×10^{-5}	1
Mum	Alt.1								0.03 (0.02)	0.03 (0.08)	0.49 (0.08)	0.54 (0.03)	0.05 (0.05)	0.1	0.47
Still	Alt.1								0.06 (0.03)	4×10^{-7} (2×10^{-8})	0.35 (0.02)	0.42 (0.02)	0.15 (0.06)	2×10^{-3}	1
Dead	Alt.1								0.03 (0.03)	0.14 (0.09)	0.45 (0.09)	0.62 (0.03)	0.05 (0.05)	0.09	0.08
Alive	Alt.1								0.56 (0.53)	2.84 (1.57)	7.07 (1.5)	10.46 (0.52)	0.05 (0.05)	0.09	0.05
Tof	Alt.1								0.71 (0.49)	2.77 (1.15)	5.21 (1.08)	8.69 (0.44)	0.08 (0.06)	0.03	0.01
Fmor	Alt.1								0.01 (5×10^{-3})	0.03 (0.01)	0.05 (0.01)	0.09 (4×10^{-3})	0.08 (0.05)	0.02	0.02
Gest	Alt.1								0.92 (0.31)	0.11 (0.66)	3.6 (0.65)	4.63 (0.24)	0.2 (0.06)	1×10^{-4}	0.65

Trait	Model	σ^2_G	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P	σ^2_G	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum	Alt.2	4×10^{-3} (2×10^{-3})	4×10^{-4} (2×10^{-3})	0.1 (3×10^{-3})	0.1 (2×10^{-3})	0.04 (0.02)	4×10^{-4}	0.54	0.02 (0.02)	0.01 (0.08)	0.44 (0.08)	0.47 (0.02)	0.04 (0.04)	0.16	0.66
Still	Alt.2	5×10^{-3} (3×10^{-3})	5×10^{-3} (4×10^{-3})	0.19 (0.01)	0.2 (5×10^{-3})	0.02 (0.02)	0.04	0.09	0.02 (0.02)	0.03 (0.07)	0.3 (0.06)	0.36 (0.02)	0.07 (0.06)	0.09	0.47
Dead	Alt.2	0.01 (4×10^{-3})	0.01 (4×10^{-3})	0.22 (0.01)	0.24 (0.01)	0.02 (0.02)	0.04	0.03	0.01 (0.02)	0.15 (0.08)	0.36 (0.08)	0.52 (0.02)	0.02 (0.04)	0.42	0.08
Alive	Alt.2	0.03 (0.02)	0.04 (0.02)	1.24 (0.03)	1.31 (0.03)	0.02 (0.02)	0.06	0.03	0.23 (0.42)	2.32 (1.44)	6.22 (1.39)	8.77 (0.43)	0.03 (0.05)	0.3	0.11
Tof	Alt.2	0.7 (0.21)	0.76 (0.2)	6.73 (0.18)	8.19 (0.21)	0.09 (0.03)	7×10^{-6}	2×10^{-5}	0.73 (0.49)	2.54 (1.18)	5.31 (1.1)	8.58 (0.43)	0.08 (0.06)	0.03	0.02
Fmor	Alt.2	1×10^{-4} (1×10^{-4})	4×10^{-4} (2×10^{-4})	0.01 (2×10^{-4})	0.01 (2×10^{-4})	0.01 (0.01)	0.14	0.01	3×10^{-3} (3×10^{-3})	0.03 (0.01)	0.04 (0.01)	0.07 (4×10^{-3})	0.04 (0.05)	0.15	0.02
Gest	Alt.2	0.54 (0.09)	0.47 (0.06)	1.16 (0.03)	2.18 (0.07)	0.25 (0.04)	$<3\times10^{-16}$	$<3\times10^{-16}$	0.92 (0.31)	0.17 (0.66)	3.46 (0.66)	4.55 (0.24)	0.2 (0.06)	8×10^{-5}	0.59

¥Transformed log(trait + 1).†Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

In epidemic phase, basic model, all traits show a significant (P<0.05) heritable genetic effect except for *Mum* where an effect is indicated at P=0.06. When the epidemic effect is taken into account in the model (Alt.1), a reduction in the size of the additive genetic variance is seen with corresponding p-values only significant at P<0.1.

No significant heritable genetic effect greater than zero is detected in the model where the dynamic epidemic trend (Alt.2) is taken into account.

A heritable genetic effect is detected in the joint analysis for the Basic and Alt.1 models. While improvements are seen in the power of the analysis, as compared to individual farm 1 estimates, the size of the additive genetic variance estimate is limited by poor (<2) estimate to standard error ratios for some traits. In the basic model traits showing higher estimate to standard error ratio are seen for *Still* (2.98), *Tof* (1.81) *Fmor* (1.83) and *Gest* (3.08).

In the joint analysis consideration was also made for the incorporation of the unknown epidemic. In this analysis records were available for 1,031 litters from 934 sows with 97 sows having repeated measures. Estimates of variance components and heritability along with the LRT P-values of the random effects are shown in Table 3.9. Data from the single record analysis of Joint Farm data is shown in the appendix Table A.10.

Table 3.9 – Joint Farm Variance Components and Heritability Estimated Using the G Matrix with Multiple Record Data Epidemic Including Unknown Phase

Trait	Model	σ^2_G	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum	Basic	0.03 (0.02)	0.03 (0.06)	0.48 (0.06)	0.53 (0.02)	0.06 (0.04)	0.04	0.46
Still	Basic	0.07 (0.02)	5×10^{-8} (3×10^{-9})	0.35 (0.02)	0.42 (0.02)	0.16 (0.05)	6×10^{-4}	1
Dead	Basic	0.04 (0.03)	0.12 (0.07)	0.47 (0.06)	0.63 (0.03)	0.07 (0.05)	0.05	0.05
Alive	Basic	0.76 (0.52)	3.01 (1.1)	6.53 (0.99)	10.3 (0.49)	0.07 (0.05)	0.03	0.01
Tof	Basic	0.88 (0.48)	1.16 (0.91)	6.77 (0.87)	8.81 (0.42)	0.1 (0.05)	0.01	0.09
Fmor	Basic	0.01 (5×10^{-3})	0.03 (0.01)	0.05 (0.01)	0.09 (4×10^{-3})	0.1 (0.05)	4×10^{-3}	1×10^{-3}
Gest	Basic	0.81 (0.28)	0.04 (0.47)	3.62 (0.48)	4.46 (0.22)	0.18 (0.06)	2×10^{-4}	1
Mum	Alt.1	0.03 (0.02)	3×10^{-7} (2×10^{-8})	0.49 (0.03)	0.52 (0.02)	0.05 (0.04)	0.06	1
Still	Alt.1	0.05 (0.02)	2×10^{-7} (1×10^{-8})	0.36 (0.02)	0.41 (0.02)	0.13 (0.05)	2×10^{-3}	1
Dead	Alt.1	0.03 (0.03)	0.07 (0.07)	0.5 (0.07)	0.6 (0.03)	0.06 (0.05)	0.07	0.18
Alive	Alt.1	0.55 (0.48)	1.67 (1.2)	7.69 (1.16)	9.92 (0.47)	0.06 (0.05)	0.07	0.13
Tof	Alt.1	0.71 (0.47)	1.13 (0.93)	6.88 (0.89)	8.72 (0.41)	0.08 (0.05)	0.03	0.1
Fmor	Alt.1	0.01 (4×10^{-3})	0.03 (0.01)	0.05 (0.01)	0.08 (4×10^{-3})	0.09 (0.05)	0.01	0.01
Gest	Alt.1	0.78 (0.28)	0.5 (0.47)	3.09 (0.44)	4.38 (0.21)	0.18 (0.06)	3×10^{-4}	0.17
Mum	Alt.2	0.02 (0.02)	8×10^{-8} (4×10^{-9})	0.43 (0.02)	0.45 (0.02)	0.04 (0.04)	0.13	1
Still	Alt.2	0.02 (0.02)	0.04 (0.04)	0.3 (0.04)	0.35 (0.02)	0.05 (0.05)	0.18	0.2
Dead	Alt.2	0.01 (0.02)	0.05 (0.06)	0.44 (0.06)	0.5 (0.02)	0.02 (0.04)	0.3	0.25
Alive	Alt.2	0.24 (0.39)	0.79 (1.11)	7.31 (1.09)	8.34 (0.39)	0.03 (0.05)	0.26	0.34
Tof	Alt.2	0.72 (0.47)	0.96 (0.92)	6.94 (0.89)	8.62 (0.41)	0.08 (0.05)	0.02	0.15
Fmor	Alt.2	3×10^{-3} (3×10^{-3})	0.02 (0.01)	0.04 (0.01)	0.07 (3×10^{-3})	0.05 (0.04)	0.11	0.04
Gest	Alt.2	0.77 (0.27)	0.52 (0.46)	2.97 (0.43)	4.27 (0.21)	0.18 (0.06)	2×10^{-4}	0.17

¥Transformed log(trait + 1). - denotes inestimable.

As shown in Table 3.9, all reproductive traits show a significant ($p<0.05$) additive genetic variance with lower standard errors than in the analysis excluding the unknown epidemic. When the epidemic effect is taken into account in the Alt.1 model, all traits indicate an effect at $P<0.1$, with *Still Tof* and *Fmor* significant at $P<0.05$). Only *Tof* and *Fmor* traits show a significant ($P<0.05$) heritable effect when the dynamic epidemic effect is accounted for in Alt.2.

Fewer traits show a significant additive genetic effect in the single record analysis than using multiple records per animal. In epidemic phase basic models, the disease traits *Still* and *Fmor,* and the non-disease traits, *Tof* and *Gest* show a significant ($P<0.05$) additive genetic

variance. Other disease traits (*Mum, Dead* and *Alive*) indicate an effect at $P < 0.1$, with estimates slightly greater than their standard error.

3.04 Discussion

3.04.1 Summary of Main Findings

Heritability and variance components have been estimated in the available data. Across the analyses heritability estimates show an increase in epidemic phase across all the disease indicator traits (*Mum, Still, Dead Alive* and *Fmor*) identified in chapter 2. These increases occur as a result of a proportionally greater increase in additive genetic variance (σ^2_A) than phenotypic variance. The estimates themselves vary between analysis and model.

In the Farm 1 analyses the highest heritability estimates were seen under the basic model using the A matrix. The highest reported value was for the *Wean* trait at 0.41 (s.e. 0.08) in the epidemic phase. The remaining disease indicator trait using the basic model with the A matrix (*Mum, Still Dead Alive* and *Fmor*) ranging from 0.16 (s.e. 0.06) to 0.14 (s.e. 0.06) in epidemic phase. These low/moderate estimates contrast with very low estimates for disease indicator traits in the non-epidemic phase between 0.02 (s.e. 0.01) and 0.04 (s.e. 0.01).

Using the G matrix many estimates showed a slight reduction in both additive genetic variance and heritability from those estimated using the A matrix for most traits (*Mum, Dead, Alive*, *Tof*, and *Fmor*). The *Still* trait however shows a consistent estimate using both the A and G Matrix under both the Basic ($h^2 = 0.16$) and Alt.1 ($h^2 \sim 0.09$) models, for which a significant h^2 estimate was obtained.

The discrepancies in h^2 estimates between the G and A matrix for some traits could be the result of a couple of factors. 'Missing heritability' has been the source of much research in the literature (reviewed in Manolio *et al.,* (2009)). Some of the explanations for missing heritability put forward explore the potential confounding within pedigree relationships of both permanent environmental effects and sampling bias with respect to non-additive genetic

and additive genetic interactions. It is also possible that the assumptions included in the calculation of the A matrix are more robust to the population architecture (Amin *et al.,* 2007).

By fitting the A and G matrix simultaneously evidence was shown of substantial confounding between the A matrix and the G matrix. The A matrix absorbing a large proportion of the additive genetic variance. Only the *Wean* trait in Epidemic Phase and the *Tof* trait in non-epidemic phase showing a significant effect under in both additive genetic components, in both cases most of the variance was captured by the A matrix.

Across all Farm 1 analyses the alternative models reported lower additive genetic variances, and as a result lower heritabilities than the basic models. In most cases under alternative model 2, in epidemic phase, additive genetic variance was inestimable. Interestingly, in the non-epidemic phase when trait trends are fitted no impact on heritability and additive genetic variance estimates was observed. This either suggests that in epidemic phase the epidemic effect not only captures environmental differences in exposure, but absorbs genetic effects or it could highlight some inflation of the basic model estimates by cohort effects.

In farm 2, with considerably fewer animals, where only the *Still* trait indicated ($p=0.06$) a significant additive genetic variance in epidemic phase providing an heritability estimate of 0.18 (0.14). The low numbers involved providing insufficient power to detect a heritable effect in all traits.

Proportionally greater increases are seen in the genetic component as compared to the corresponding standard error for farm 2 using alternative model 2 as compared to the basic model across traits. While an additive genetic variance was only indicated at $P<0.08$ for the *Still Alive* and *Fmor* traits, this pattern was repeated in non-epidemic phase with all traits showing the improvement. This contrasts Farm estimates 1 where under Alternative model 2 a reduction in additive genetic variance, additive genetic component to standard error ratio

and heritability are seen as compared to the basic model. While accounting for the magnitude of epidemic severity is intuitively required when estimating additive genetic variance, it would appear that on farm 1, either the genetic component is confounded in epidemic ID or that a reduction in power to estimate effects limits the calculation of an additive genetic component in this model.

The different response of the two farms to alternative model 2 is interesting. One of the most noticeable difference between the animals on farm 1 and the animals on farm 2 is breed diversity. It may be that using a cohort based estimate of disease outcome confounds more of the genetic effect in the more diverse population than in a more homogenous one. A downward bias of heritability estimation has been shown for resilience type traits if by fitting cohort estimates of disease burden you confound resistance (Doeschl-Wilson *et al.,* 2012).

In the joint farm analysis, a statistically significant heritable component was observed for all traits ($p<0.05$) except for *Mum* in which an effect was indicated at $p=0.06$. These h^2 estimates are similar to the estimates generated for the farm 1 analysis. Low power limits the conclusions which can be made regarding the contribution of the farm 2 h^2 estimates which were generally higher. Farm 1 made a greater contribution in terms of number of animals than farm 2 with 605 and 329 sows incorporated into the genetic analysis respectively.

Evidence was seen for increased power in the joint analysis, as compared to the individual farm analyses for all traits (except *Alive*) and models, as proportionally greater increases in additive genetic variance estimate than corresponding standard errors. The most considerable improvements to power were seen in the joint, repeated measures analysis, where data including the unknown epidemic was used. An increase in the additive genetic variance to standard error ratio is observed across traits and models as compared to both the joint analysis and the individual Farm 2 analysis where the unknown epidemic is included.

Across farms and models, a consistent heritability of ~0.17 was estimated for the *Still* trai While in farm 1 this dropped slightly in alternative model 1 and was inestimable in model 2, this was consistent with the pattern observed for other traits between models in this farm.

The genetic variance for *Fmor,* was also broadly significant across analyses using the Basic and Alternative model 1. Exceptions to this were for Farm 1, G matrix, alternative model 1 (p=0.09) and the basic model for Farm 2, though here a heritable effect was indicated under alternative model 2. While raised in the farm 1 a matrix estimates, heritability for *Fmor* was fairly consistent at ~0.1 under the basic model reducing to ~0.07 under alternative model 1 where an effect is indicated, this increased to 0.167 (p=0.074) for alternative model 2 for farm 2. In the joint farm analysis, it remained fairly consistent between basic and alternative model 1, though a slight reduction was seen between multiple record and single record analyses with estimates of ~0.075 and ~0.11 respectively.

3.04.2 Comparisons with Previous Studies

As discussed in the introduction differences in model limit direct comparison with previously published literature given the models used here account for *Tof* as a fixed covariate. Despite this difference very similar heritability estimates are generated here for *Mum, Still, Dead, Alive* and *Tof* under comparable A-matrix analysis using the basic model as were observed in (Lewis *et al.,* 2009b). A considerably greater heritability is observed for the *Wean* trait (0.407 s.e. 0.08 *cf.* 0.15 s.e 0.03). This is probably due to a slightly modified definition of disease phase and the changes made to the model. For example, in this study the model for the *Wean* trait additionally includes the interaction between the fostering numbers (Net Fostered) and total litter size (*Tof*). While the risks associated with fostering have been demonstrated in the literature (Mccaw, 2000) intuitively, risk of pre-weaning losses will occur as a function of both the litter size (*Tof*) and the numbers of animals fostered, as measures of the relative risks posed by the numbers in each group.

Non-epidemic phase *Alive* trait estimates are considerably lower than those reported in the literature at 0.07-0.11 (Holm *et al.,* 2004; Serenius *et al.,* 2004; Lewis *et al.,* 2009b; Vidovic *et al.,* 2012; Sevón-Aimonen & Uimari, 2013; Serão *et al.,* 2014). This was attributed to the effect of fitting the Alive trait conditional on the *Tof* as a covariate. When the *Tof* covariate was dropped from the model the heritability for *Alive* increased to ~0.075 across both farms and in the joint analysis (±0.005), consistent with previous estimates. Low non-epidemic phase estimates are consistent with those published in the literature for *Still* when litter size is included as a covariate (Knoll 2002). Estimates of heritability presented for *Tof* itself at 0.07-0.14 are consistent too with previously reported published average of 0.10 (Rothschild & Ruvinsky, 2011).

It is worth noting too that in the previous study the permanent environmental effect, while originally fitted, was dropped as a result of a lack of significance. This was observed for many traits here most likely a result of a lack of repeated records given the timescales involved in epidemic duration, rather than a lack of a permanent environmental effect. For the *Dead Alive* and *Fmor* traits a statistically significant permanent environmental variance was demonstrated in these data though could not be shown for *Mum* or *Still* possibly owing to narrower phenotypic variance. Retaining the permanent environmental effect, regardless of the demonstration of a significant variance component, guards against this potential source of bias in the estimation of the additive genetic variance.

The use of the counts of the various reproductive outcomes as a quantitative trait was favoured here as compared to the binary analysis, used as a resistance type trait in the previous study. Using the arbitrary threshold employed in Lewis *et al.,* (2009b) to convert the quantitative trait to a binary outcome exposes the analysis to inaccuracies of classification of animals as diseased or healthy discussed in the previous chapter.

The heritability estimates generated by (Serão *et al.,* 2014) for an independent dataset used a similar model to that described in alternative model 2 (Alt.2), though omits litter size as a

covariate. Whilst here farm 1 data failed to generate estimates under the Alt.2 Model, the additive genetic variance estimated on farm 2 appear more robust to the fitting of trait trend which generated indicative (p= 0.06-0.08) results for *Still* (h^2=0.19 s.e. 0.15) *Alive* (h^2=0.17 s.e.0.14) and *Fmor* (h^2=0.17 s.e. 0.14). These are slightly greater than the estimates in Serão *et al.,* (2014). They however are more in keeping with the approximate Basic model estimates on this farm, and the A matrix estimates generated under the basic model for farm 1. It is also worth noting that the non-epidemic ("Pre-PRRS Phase") heritability estimates presented in Serão et al., (2014) are considerably greater than those presented here. This is attributed to the effect of including *Tof* as a covariate.

The estimates provided in Serão et al. 2014 do not show a h^2 increase in disease phase the *Mum, Still* and *Fmor* traits as compared to non-epidemic phase. Heritability estimate to standard error ratios are low for these traits (<1.16) indicating potential issues of power. Given that σ^2_A values were not provided in Serão *et al.,* (2014) it is not possible to see whether the reductions seen in epidemic phase heritability occur due as a reduction in σ^2_A or as a result of proportionately greater increase in σ^2_P than σ^2_A; the latter would still indicate a genetic component in response to PRRSV, additional to an effect in the absence of PRRSV. For the remaining traits in Serão *et al.,* (2014), comparable to those used in this research estimates are presented for *Mum* (h^2=0.08), *Alive* (h^2=0.09) and *Dead* (h^2=0.12), which generally concur with the order of magnitude of the estimates presented here except for the *Wean* trait where the same differences in model are noted as with the Lewis *et al.,* (2009b) study.

Genetic variance was demonstrated too for reproductive performance traits in data including the unknown epidemic. In the joint analysis, comparing the additive genetic variance and standard error between the confirmed epidemic and analysis where the unknown epidemic is included a small decrease is seen in estimate, though a proportionally greater decrease is

seen in the standard error suggesting improvements in power. In these data the unknown epidemic represented 23% of the litters.

3.04.3 Further Work

This study was restricted to univariate analysis. Further work could explore multi-variate analyses to explore the genetic correlations between traits. A recent study by Lough *et al.,* (2017) suggests using a multivariate genetic covariance analysis that a performance in the absence of infection (body weight), resistance (control of virus replication) and performance under infection (growth rate under challenge), show a strong positive genetic correlation (also reported in Hess *et al.,* (2016) and Boddicker *et al.,* (2012)). However, the correlation was significantly different from 1, suggesting that the variation in resilience is not fully accounted for by genetic variance in resistance or performance in the absence of infection. These analyses provide useful insights into the complex relationships of the components of resilience.

Some success has been achieved in looking at the genetic basis of variance from distant populations (Hayes et al., 2009c; Riggio et al., 2014; Hess et al., 2016). A method applied in these studies considers separate populations as distinct in the G matrix by setting relationships between separate populations to zero. Given relationships described using pedigree data for Farm 1 give considerably larger variance estimates than the relationships estimated using SNP data it is possible that population differences impede the ability of the G matrix to accurately represent IBD. If this is the case then it may be possible to modify the G matrix such that it better represents IBD with a view to capturing more of the variance that is demonstrated using the A matrix. This will be explored further in later chapters.

The results from the analysis including the unknown epidemic suggest that the threshold-threshold method employed in Lewis *et al.,* (2009a) could be used to identify epidemics on the basis common clinical signs (independent of aetiology), which can then be used to test

for a common genetic "robustness to increases in reproductive failure" across aetiologies. Exploiting a common outcome to look for genetic merit against multiple pathological causes. While it is improbable that the unknown epidemic is PRRSV, further work using confirmed epidemics of different pathogens would be needed to formally establish this, given the unknown cause of epidemic 5.

3.04.4 Conclusion

A heritable signal has been demonstrated for reproductive performance in these data. These broadly concur with estimates of similar studies, though a considerably higher heritability is reported for the *Wean* trait. The models used in this analysis were designed to account for environmental sources of variation of reproductive outcomes with specific reference to population structure and permanent environmental effects. For all disease indicator traits where a significant heritable effect is reported show an increase in additive genetic component with regards to the same component in data where PRRSV is not detected. This demonstrates that selection methods can be employed to improve the host with regards reproductive success under PRRSV challenge.

Epidemiological factors were also considered using alternative models designed as part of this research. The reduction in genetic effect using the alternative models, seen in Farm 1 and the joint analysis could suggest confounding of the genetic effect. Given a lack of additional data it is not possible to explore this further. Whilst estimates of heritability are congruent with previously reported estimates to account for any bias which could occur by accounting for variation between and within epidemic.

Combining the datasets in a joint analysis improved statistical power, in the joint analysis a significant additive genetic variance was demonstrated across all disease indicator traits. Similarly, including data from an unknown epidemic as a result of a potentially different agent also seemed to improve power.

Results suggest that including the *Tof* as a covariate creates a greater contrast between non-epidemic and epidemic heritability estimates. This may provide a useful signal for investigating the distribution of PRRS related mortality throughout the pig genome.

Chapter 4. Population Structure and the Genomic Relationship Matrix

4.01 Introduction

As observed in Chapter 3 there were considerable differences in heritability estimates for Farm 1 between estimates from relationship matrices derived using pedigree (A matrix) and genomic (G matrix) methods. Consistently higher estimates were seen using the A matrix with the basic model except for the Still trait, where no difference was observed. Several studies have found additive genetic variance estimates using pedigree information higher than that estimated using genomic relationships (e.g. Veerkamp *et al.,* (2011) in cattle and Forni *et al.,* (2011) in pigs). It has been suggested that an upward bias in σ^2_A estimates, obtained with the A matrix can occur from the cofounding of common environment within additive relationships (Vinkhuyzen *et al.,* 2013). A further limitation to the use of pedigrees is that they treat the founders as unrelated which may or may not be the case and rely solely on the explicit relationships described (Speed & Balding, 2014).

Genomic relationship matrices (GRMs) are more accurate in describing identity-by-descent, given the ability to account for Mendelian sampling and for relationships not explicitly detailed in a pedigree (Hill & Weir, 2011). Despite the advantages to the use of GRMs, some studies have however suggested in the presence of population stratification that relationships between distant population maybe misleading with regards to contribution to the genetic effects (Hayes *et al.,* 2009a). As such whilst G matrix estimates may represent a more accurate reflection of identity by descent, the use of identity-by-state can infer relationships which may be misleading in terms of estimating genetic effects

To overcome this several studies have used a prior that distant populations are unrelated, fixing the value of the block diagonal relationships between populations (Hayes *et al.,* 2009a; Riggio *et al.,* 2014; Hess *et al.,* 2016). Indeed applying GBLUP to predict milk traits across cattle breeds, a greater accuracy was obtained using GRM modified in this way (Hayes *et al.,* 2009a). This method was also used in Hess *et al.,* (2016) to consider animals of different genetic backgrounds in experimental PRRSV trials. Estimates from the original

GRM were contrasted with the modified GRM to investigate how the genetic factors between animals of different genetic backgrounds influenced the variation in different traits. For some traits, differences in the estimates from two forms of GRM indicated differences in the contribution of the genetic factors between genetic backgrounds on trait variance (Hess *et al.,* 2016). Additionally this method has also been incorporated into regional heritability mapping (a methodology for the localisation of regional effects, Nagamine et al. (2012)); Riggio *et al.,* (2014) fixes the between population relationship to 0, to consider linkage phases discrete in different populations.

The GRM calculations shown in section 3.02.2 correct for allele frequency and observed *vs.* expected homozygosity. This incorporates the assumption that allele frequency in the observed population is representative of the allele frequency in a base population, to which all members of the observed population belong (Powell *et al.,* 2010). This cannot be said if sub-populations within the observed population are reproductively isolated, and the allele frequencies are different in these sub-populations.

Issues with regards to population structure may be true of the breeds observed in farm 1 (see Figure 3.3), where several breeding lines show very low similarity and are distinct within the pedigree (Figure 3.2). These differences are also seen in joint farm analysis where farm 1 makes a greater contribution in terms of the number of litter records than farm 2. Given the existence of cross breeds in the data some of these relationships were maintained in this analysis (see materials and methods).

This chapter investigates how heritability estimates differ using different prior assumptions on the relationships between populations. In particular; the assumption that the between population (breed and, where applicable, farm) relationships are uninformative with regard to the estimation of additive genetic variance is tested. Two alternative methods are used create the modified G matrix under this assumption:

Method One

This method used in Riggio et al. 2014 and Hess et al. 2016, considers populations and farms as discrete by setting the block diagonal relationship between distinct breeds to zero in a G matrix calculated on the whole population.

Method Two

This method uses knowledge of the breeds and crosses to create separate G matrices calculated using only reference populations pertinent to the relationships being estimated,

then combines these matrices to only describe known within breed and between breed relationships in the population. This idea is shown graphically in Figure 4.1.

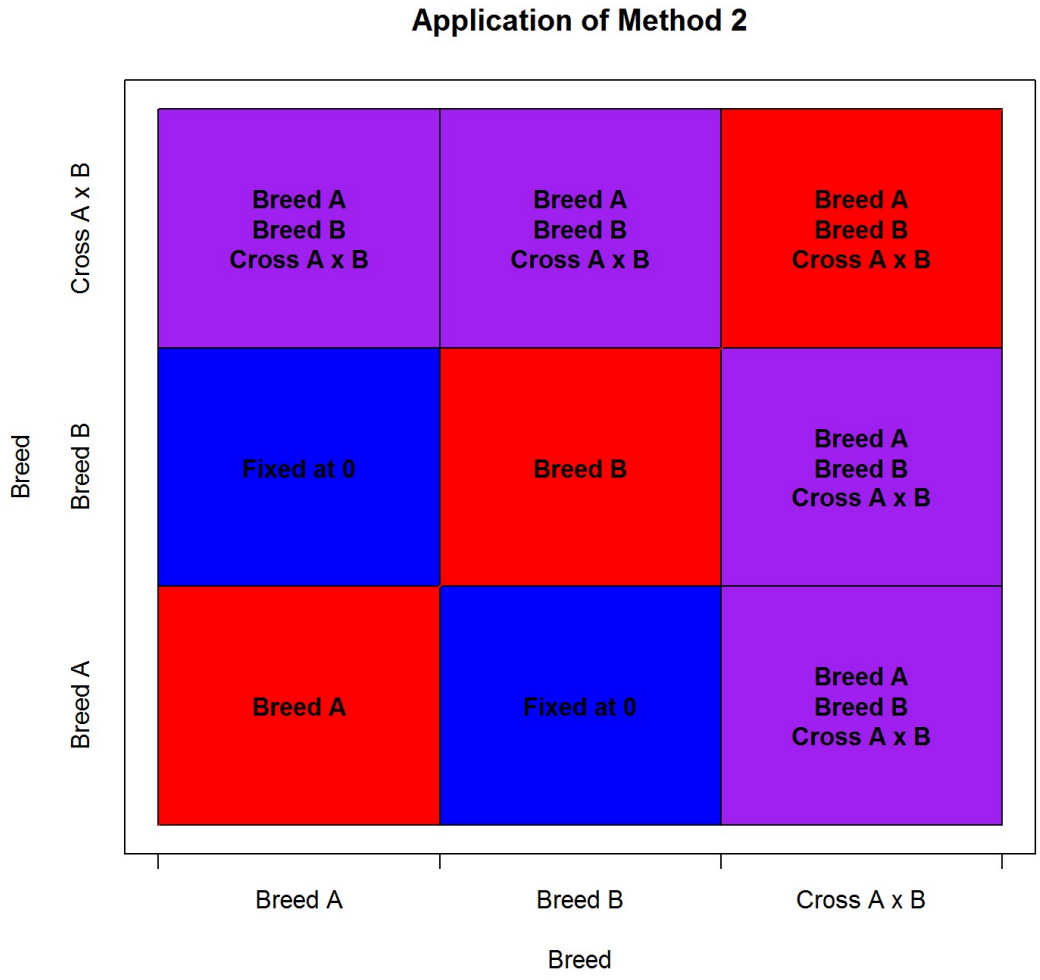

Figure 4.1 – Application of Method Two

Application of method two for three separate lines, two discrete breeds (A and B) and an A × B cross. Intersections detail the population used in the calculation of a sub-matrix used to populate each portion of the G matrix.

These methods will be explored using the Farm 1 data, for which pedigree and genomic data is available. Modified genomic relationship matrices indicating improved power to estimate the additive genetic variance, will be reproduced in the joint farm analysis.

4.02 Materials and Methods

Variance component analyses are as described in section 3.02.3 using ASREML 3.0. The data used is described for both farms in section 2.02.1. Data was partitioned according to an adapted threshold-threshold method described in (Lewis *et al.,* 2009a) into epidemic and non-epidemic phases described in section 2.02.2. In this analysis, the epidemic phase is analysed, for which a significant heritable component demonstrated in chapter 3. Farm 1 is used to consider the difference between and additive genetic variance estimated using pedigree information, the full genomic relationship matrix and that modified using the two methods. The data from the combined farms in then also considered.

Models and traits are described in 3.02.3. Only the Basic model and alternative model 1 (Alt.1) are considered here. Compared to the basic model, alternative model 1 fitted additionally a unique identifier for each epidemic, to account for differences between epidemic severities. In the joint analysis given the hierarchical structure, Epidemic ID was nested within Farm.

4.02.1 Genomic Relationship Matrices

The calculation of the full genomic relationship matrix (GRM) is shown in section 3.02.2. In method 1 the between breed relationships are fixed at zero in the full GRM. Relationships between crosses and their founding breeds were maintained. Here the between breed variation is removed from the analysis whilst retaining relationships within the same breed and with the crosses.

Method 2 investigates if combing separate sub-G-matrices on the separate subpopulations could be used as an alternative method for the creation of an IBS containing discrete populations. For each breed only those animals within that breed were used for calculating the within breed relationships. For a crossbred line, animals from the founding breeds and the cross were used for calculating the relationship between the cross and its founding breeds

and within the cross, but not inferring any relationships within the founding breeds; the relationship between breeds was fixed at 0.

Genomic relationship matrices were calculated in R using GenABEL (Aulchenko *et al.,* 2007b).

4.03 Results

4.03.1 Farm 1

Given the considerable reduction in the number of animals for which genetic data is available as compared to pedigree data (994 *cf.* 606 in epidemic phase data), additive genetic variance estimates using pedigree information were recalculated using only animals for which genetic data was available. A comparison of the Farm 1 relationship matrices is shown in Figure 4.2.

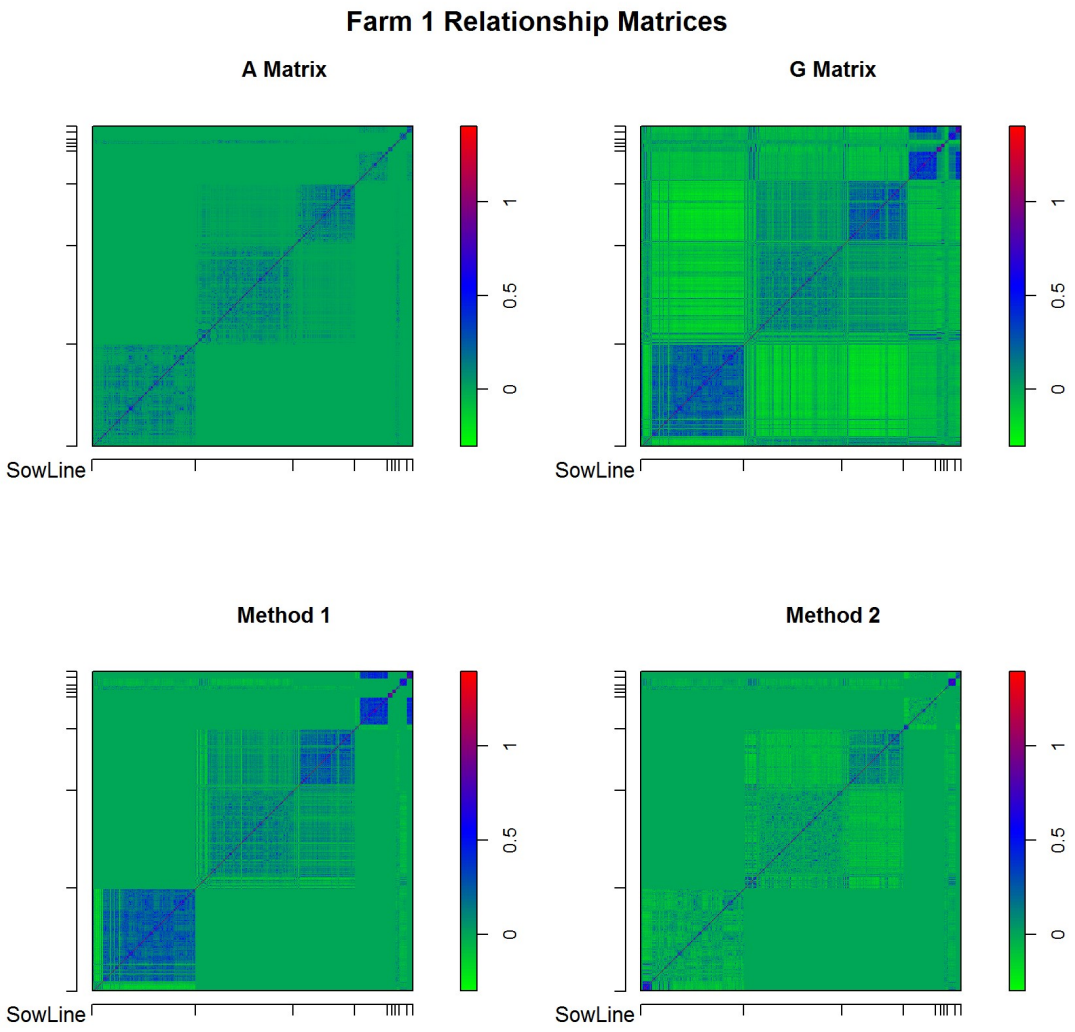

Figure 4.2 – Relationship Matrices Comparing A Matrix, G Matrices, Including Full, Calculated by Breed and Block Diagonal Between Breeds Fixed at Zero

Heat plot showing relationship matrices showing the additive relationship calculated using the pedigree data (A Matrix), the full genomic relationship matric (G matrix), a genomic relationship matrix with the between breed relationships fixed at zero (Method 1) and a single matrix compiled from G Matrices calculated separately by breed (Method 2).

Whilst similar patterns are observed in all four matrices some differences can be observed. Generally closer genomic relationships are observed when the whole population is used to estimate genomic relationships than when the matrix is compiled from separate sub-matrices This appears more congruent with the A matrix figures which are based on direct relationship records. This suggests that the matrix calculated under method 2 is not an

accurate reflection of IBD, areas showing high (>0.25) levels of relatedness in the A matrix are not reflected in the method 2 matrix.

Variance components estimated using the four different matrices on the same single record farm 1 epidemic data, are shown in Table 4.1.

Table 4.1 – Farm 1 Epidemic Phase Variance Component Estimates Using the Different Matrices

Trait	Model	A Matrix			G Matrix			GRM Fixed at Zero by Breed			GRM Calculated Separately by Breed		
		σ^2_A (SE)	h^2 (SE)	σ^2_A LRT	σ^2_A (SE)	h^2 (SE)	σ^2_A LRT	σ^2_A (SE)	h^2 (SE)	σ^2_A LRT	σ^2_A (SE)	h^2 (SE)	σ^2_A LRT
Mum¥	Basic	0.09 (0.06)	0.12 (0.08)	0.02	0.04 (0.04)	0.06 (0.06)	0.09	0.04 (0.04)	0.05 (0.06)	0.17	0.02 (0.04)	0.03 (0.05)	0.27
Still¥	Basic	0.13 (0.05)	0.27 (0.1)	4×10^{-4}	0.08 (0.04)	0.16 (0.07)	<0.001	0.09 (0.04)	0.2 (0.08)	<0.001	0.08 (0.03)	0.18 (0.07)	<0.001
Dead¥	Basic	0.14 (0.07)	0.2 (0.1)	7×10^{-3}	0.05 (0.04)	0.07 (0.06)	0.08	0.05 (0.05)	0.07 (0.07)	0.13	0.04 (0.04)	0.05 (0.06)	0.17
Alive	Basic	3.22 (1.52)	0.23 (0.1)	0.002	1.26 (0.93)	0.09 (0.06)	0.03	1.29 (1.06)	0.09 (0.07)	0.07	0.76 (0.88)	0.05 (0.06)	0.17
Tof	Basic	1.77 (0.9)	0.2 (0.1)	0.006	0.93 (0.6)	0.11 (0.07)	0.02	1.3 (0.7)	0.15 (0.08)	<0.001	1.63 (0.69)	0.19 (0.08)	0.002
Fmor	Basic	0.03 (0.01)	0.24 (0.1)	0.002	0.01 (8×10^{-3})	0.1 (0.06)	0.02	0.01 (9×10^{-3})	0.1 (0.07)	0.04	8×10^{-3} (7×10^{-3})	0.07 (0.06)	0.11
Wean	Basic	4.71 (1.27)	0.48 (0.11)	<0.001	2.98 (0.88)	0.3 (0.08)	<0.001	3.7 (0.96)	0.38 (0.08)	<0.001	2.63 (0.67)	0.28 (0.06)	<0.001
Gest	Basic	1.84 (0.65)	0.32 (0.1)	<0.001	1.5 (0.53)	0.26 (0.08)	<0.001	1.34 (0.53)	0.23 (0.08)	<0.001	0.85 (0.41)	0.15 (0.07)	0.002
Mum¥	Alt.1	0.07 (0.05)	0.1 (0.08)	0.05	0.03 (0.04)	0.05 (0.06)	0.15	0.02 (0.04)	0.03 (0.06)	0.29	9×10^{-3} (0.03)	0.01 (0.05)	0.51
Still¥	Alt.1	0.09 (0.04)	0.21 (0.1)	<0.001	0.05 (0.03)	0.1 (0.07)	0.03	0.06 (0.03)	0.13 (0.08)	0.02	0.05 (0.03)	0.12 (0.07)	0.03
Dead¥	Alt.1	0.09 (0.06)	0.13 (0.09)	0.04	0.01 (0.04)	0.02 (0.05)	0.42	8×10^{-3} (0.04)	0.01 (0.06)	0.58	-	-	-
Alive	Alt.1	2.13 (1.36)	0.16 (0.1)	0.02	0.57 (0.82)	0.04 (0.06)	0.21	0.47 (0.91)	0.04 (0.07)	0.33	-	-	-
Tof	Alt.1	1.3 (0.88)	0.15 (0.1)	0.04	0.72 (0.57)	0.08 (0.06)	0.05	0.96 (0.68)	0.11 (0.08)	0.05	1.31 (0.69)	0.15 (0.08)	0.01
Fmor	Alt.1	0.02 (0.01)	0.17 (0.1)	0.02	5×10^{-3} (7×10^{-3})	0.05 (0.06)	0.17	5×10^{-3} (8×10^{-3})	0.04 (0.06)	0.27	1×10^{-3} (6×10^{-3})	0.01 (0.05)	0.6
Wean	Alt.1	1.49 (0.86)	0.18 (0.1)	0.03	0.99 (0.65)	0.12 (0.08)	0.04	1.54 (0.73)	0.19 (0.08)	<0.001	1.23 (0.61)	0.15 (0.07)	<0.001
Gest	Alt.1	1.55 (0.62)	0.28 (0.1)	<0.001	1.29 (0.51)	0.22 (0.08)	<0.001	1.12 (0.51)	0.2 (0.08)	0.004	0.65 (0.4)	0.12 (0.07)	0.02

¥Transformed log(trait + 1). An effect on σ^2_P was not observed and so is not shown.

Variance components and heritability estimated using the A matrix, when the data is limited to just animals with genetic data, show similar patterns as observed using the all animals for which pedigree data is available in Chapter 3. Whilst differences are observed in the estimates of σ^2_A and h^2 when the number of animals is limited, no systematic difference is observed. An expected increase in standard error and reduction in P values is seen due to the reduction in the number of records. These suggest that the systematic reduction seen in G matrix estimates is not simply a function of a reduction in power owing to less records. The A matrix variance component and heritability estimates are still considerably greater than those generated using the G matrix. No effect observed on σ^2_P suggesting that the changes in h^2 are as a result of differences in σ^2_A.

When the between breed relationships are fixed at zero (method 1), a slight increase in heritability is observed in some traits (*Still, Dead, Alive, Tof, Fmor* and *Wean*) as compared to the estimates generated under the full G matrix. These increases are small, with heritability estimated in the same order of magnitude. Increases are seen too in the standard error estimates of the *Mum* and *Dead* additive genetic effect, which are no longer significant at $p < 0.1$. For the *Still, Tof*, *Fmor* and *Wean* traits, a proportionally greater increase is seen in estimate than standard error which may indicate some improvement in power. In alternative model 1 estimates are considerably lower except using the GRM modified using method 1, for *Still Fmor* and *Tof* continue to show a small increase.

For results generated under method 2, many traits show a small reduction in the variance estimates and increases in standard error, with the exception of *Tof* and *Still*. A considerable impact is seen in the p values for the additive genetic effect with only *Still, Fmor Wean* and *Gest* remaining significant at $p > 0.1$. Increases are seen in the standard error and decreases in the standard error to estimate ratio. This indicates a considerable reduction in power using method 2 and a possible lack of consistency in the way that the estimates are derived.

4.03.2 Joint Analysis

Discrete populations may be considered as in the joint analysis in two ways. The relationships between farms, set to 0, or the relationships between breeds set to 0. Both applications were applied to investigate the impact on variance component estimates.

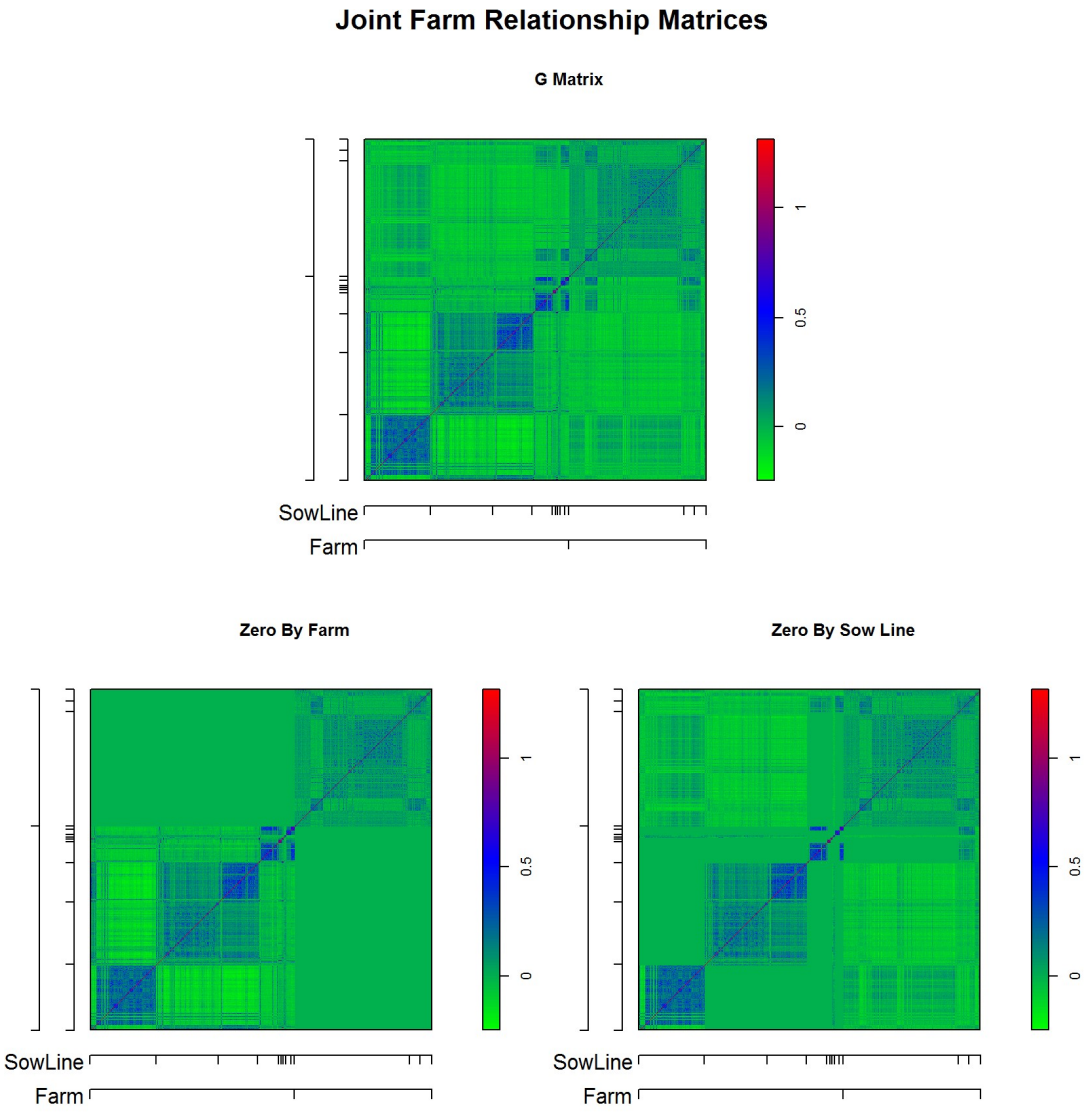

Figure 4.3 – Comparison of the Full Genomic Relationship (G matrix) and the Genomic Relationship Matrix with Between Population (Farm or Breed) Relationships Fixed at Zero

Heat plot showing relationship matrices showing the additive relationship calculated using the full genomic relationship matric (G matrix), a genomic relationship matrix with the between farm relationships fixed at zero and matrix with the between breed relationships fixed at 0.

A comparison of the estimates made using the full G matrix as compared to the G matrices with the between farm or between breed relationships fixed at 0, is shown in Table 4.2.

Table 4.2 – Variance Components and Heritability Estimated for Epidemic Phase Under the Full G Matrix and Matrices with the Cross Farm or Cross Breed Relationship Fixed at Zero

Trait	Model	Unmodified GRM				GTM Fixed at Zero Between Farm				GRM Fixed at Zero Between Breed			
		σ^2_A	σ^2_E	h^2 (SE)	σ^2_A LRT P	σ^2_A	σ^2_E	h^2 (SE)	σ^2_A LRT P	σ^2_A	σ^2_E	h^2 (SE)	σ^2_A LRT P
Mum¥	Basic	0.03 (0.03)	0.54 (0.03)	0.05 (0.05)	0.07	0.03 (0.03)	0.53 (0.03)	0.06 (0.05)	0.06	0.04 (0.03)	0.53 (0.04)	0.07 (0.06)	0.05
Still¥	Basic	0.08 (0.03)	0.35 (0.03)	0.2 (0.07)	<0.001	0.07 (0.03)	0.36 (0.03)	0.16 (0.06)	<0.001	0.08 (0.03)	0.34 (0.03)	0.2 (0.07)	<0.001
Dead¥	Basic	0.04 (0.03)	0.61 (0.04)	0.06 (0.05)	0.07	0.04 (0.03)	0.6 (0.04)	0.07 (0.05)	0.05	0.06 (0.04)	0.59 (0.04)	0.1 (0.06)	0.02
Alive	Basic	0.65 (0.58)	10.33 (0.68)	0.06 (0.05)	0.08	0.78 (0.6)	10.21 (0.69)	0.07 (0.05)	0.05	0.96 (0.69)	10.05 (0.74)	0.09 (0.06)	0.04
Tof	Basic	0.9 (0.52)	7.87 (0.55)	0.1 (0.06)	0.01	0.74 (0.52)	8 (0.56)	0.08 (0.06)	0.03	0.85 (0.6)	7.91 (0.61)	0.1 (0.07)	0.05
Fmor	Basic	0.01 (5×10^{-3})	0.09 (0.01)	0.09 (0.05)	0.01	9×10^{-3} (5×10^{-3})	0.08 (6×10^{-3})	0.1 (0.05)	<0.001	0.01 (6×10^{-3})	0.08 (6×10^{-3})	0.13 (0.06)	<0.001
Gest	Basic	1.01 (0.34)	3.71 (0.29)	0.21 (0.07)	<0.001	1.11 (0.36)	3.61 (0.3)	0.23 (0.07)	5×10^{-5}	1.02 (0.36)	3.66 (0.32)	0.22 (0.07)	1×10^{-4}
Mum¥	Alt.1	0.03 (0.03)	0.53 (0.03)	0.05 (0.05)	0.11	0.03 (0.03)	0.53 (0.03)	0.05 (0.05)	0.08	0.03 (0.03)	0.53 (0.04)	0.05 (0.05)	0.12
Still¥	Alt.1	0.07 (0.03)	0.35 (0.03)	0.17 (0.06)	<0.001	0.05 (0.03)	0.36 (0.03)	0.13 (0.06)	<0.001	0.07 (0.03)	0.35 (0.03)	0.17 (0.07)	<0.001
Dead¥	Alt.1	0.03 (0.03)	0.6 (0.04)	0.05 (0.05)	0.12	0.03 (0.03)	0.6 (0.04)	0.05 (0.05)	0.13	0.04 (0.04)	0.59 (0.04)	0.06 (0.06)	0.08
Alive	Alt.1	0.43 (0.54)	10.26 (0.66)	0.04 (0.05)	0.17	0.47 (0.55)	10.23 (0.67)	0.04 (0.05)	0.15	0.59 (0.64)	10.12 (0.71)	0.06 (0.06)	0.13
Tof	Alt.1	0.77 (0.51)	7.96 (0.55)	0.09 (0.06)	0.03	0.63 (0.51)	8.07 (0.56)	0.07 (0.06)	0.06	0.62 (0.58)	8.08 (0.61)	0.07 (0.07)	0.12
Fmor	Alt.1	0.01 (5×10^{-3})	0.08 (0.01)	0.07 (0.05)	0.03	7×10^{-3} (5×10^{-3})	0.08 (6×10^{-3})	0.07 (0.05)	0.03	9×10^{-3} (6×10^{-3})	0.08 (6×10^{-3})	0.1 (0.06)	0.02
Gest	Alt.1	0.9 (0.33)	3.75 (0.29)	0.19 (0.07)	<0.001	1.02 (0.35)	3.65 (0.3)	0.22 (0.07)	<0.001	0.92 (0.36)	3.71 (0.32)	0.2 (0.07)	<0.001

¥Transformed log(trait + 1). No effect was observed on σ^2_P which is not shown.

Slight increases are seen in the heritability estimates in the Basic Model with the between farm relationship fixed at 0, the increase in heritability outweighing the associated increase in standard error. This effect on the ratio of the σ^2_A estimate to its standard error is consistently higher when the genomic relationship is between fixed at zero between breed as opposed to farm. This suggests improved power in the ability to estimate σ^2_A when genomic relationships are fixed at zero between breeds. A slight improvement is also seen for the *Mum* trait under Alternative model 1, no effect was detected using the full G matrix however using this modified G matrix an effect is not indicated at $P<0.1$.

When the between breed relationships are fixed at Zero increases are seen in the variance component estimates for nearly all disease traits as compared to the estimates using the original G matrix. Where this effect is observed proportionally greater increase is seen in the estimate than in the standard error of the estimate. This is except for the *Still* trait. The significance of these traits also improved with all traits significant at $p<0.05$ in the basic model.

Where either the between farm or between breed relationships are fixed at zero, see an improvement in terms of lower standard error to estimate ratios, however this improvement is greater when the between breed relationships are fixed at 0,

4.04 Discussion

Using a method first explored in Hayes *et al.,* (2009a), removing the relationships between individuals in separate populations from the genomic relationship matrix; increases were seen in the ratio of the estimate to standard error ratio for the additive genetic variance. These results indicate increased power to estimate additive genetic variance, with the population structure removed from the GRM. This suggests that inferring relationships between both farm and breed could be acting as nuisance terms in the genetic analyses.

Small increases were also observed in the additive genetic variance and heritability estimates. In the farm 1 data these increases are only seen for a few traits and were small.

Using a separate method, where genomic relationship matrices were calculated separately and combined (method 2), additive genetic variance could not be accurately or consistently estimated. This was probably due to unequal representation of the breeds when calculating the allele frequencies in crosses. Methods in the literature (e.g. Makgahlela et al., 2014)) have overcome this issue by weighting the contribution to allele frequency made by founder breeds accordingly, thus compensating for unequal representation.

For the joint farm analysis increases in the additive genetic variance to standard error ratio were seen when the relationships between farm or breed were fixed at zero. The greatest increases were seen when the between breed relationships were fixed at zero. Given the relationship of effect size and standard error on power, these results suggest improved power in the ability to estimate an additive genetic effect using the method 1 modified matrix.

A similar adapted G matrix to that generated using method 1 can be used in conjunction with Regional Heritability Mapping (RHM) (Nagamine *et al.,* 2012) to explore the distribution of trait effects across the genome. This is a technique where genome data is dissected into consecutive, overlapping sections (termed windows) from which G matrices are calculated, and the significance of each window calculated. This method will be applied when applying the RHM method to the joint farm epidemic data.

Chapter 5. Genetic Associations of Reproductive Traits Under PRRSV Challenge

5.01 Introduction

A number of GWA analyses are reported in the literature exploring variation of host response to PRRSV infection, reviewed in Chapter 1. These consider a range of traits exploring variation in the growing and reproducing sow with regards to immune response, pathogen burden, weight gain and reproductive outcomes. The first GWA analysis for reproductive performance under PRRSV challenge is given in Lewis *et al.,* (2009c). Using a custom 7K SNP chip in conjunction with the farm 1 phenotype data used in this thesis. 11 SNPs were presented as significant above the genome wide significance threshold. Since these were not reported using standard unique identifiers (i.e. reference SNP numbers used in dbSNP and Ensemble, or consortium synonyms) the positions of these SNPs on the current *Sus Scrofa* genome build 10.2 could not be identified. Four SNPs are presented with mapped positions in Lewis, 2008: at SSC1 98.3 Mbp, SSC4 92 Mbp, SSC4 91Mbp and SSC14 28.5Mbp. As acknowledged in previous chapters there is a distinction in the models presented in this thesis to the ones that were applied in Lewis et al., (2019c).

Two further studies by Yang et al. (2016) and Serão et al. (2014) use reproductive traits, directly comparable to those used in this analysis. Yang et al. (2016) found 20 SNPs associated with foetal mortality in response to PRRSV with an additional three showing associations with viral load, distributed throughout the genome. For foetal mortality the individual piglet state was fitted as a binary trait, transformed using a logit link function to the underlying liability scale. Whilst litter size was included as a covariate this was in addition to the percentage of foetuses born dead in the litter, to which the individual foetus belonged. The largest SNP effect, explaining 3.36% of the total genetic variance was detected on SSC7 at 97 Mbp. A significant association was shown for number of stillborn piglets and number born dead in Serão *et al.,* (2014); located at ~32 Mbp on SSC1 explaining 11.04% and 0.81% of the total genetic variation, respectively. Again, distinctions

in the model are noted which did not include litter size, though did include the trait trend as per the Alt.2 Model in this thesis.

A summary of published associations for PRRSV related traits, regardless of trait is shown in Figure 5.1. The most relevant results for the present study are those from Lewis, (2008) presented in Lewis *et al.,* (2009c) using mortality traits (⊕); those in Yang *et al.,* (2016) using the foetal death traits (⊞); and the single association with number born dead reported in Serão *et al.,* (2014) (⊠), as these refer to reproductive performance of sows.

The best characterised and most validated genomic region associated with host response to PRRSV infection has been provided by the large scale PRRS host genetics consortium (PHGC) experiments, in which thousands of growing piglets have been infected with virulent strains of the North American PRRSV type (for review see Dekkers *et al.,* (2017)). A large effect was detected for a 0.5 Mbp region on SSC4 associated with both weight gain and cumulative serum viral load within the first 3 weeks post infection (Boddicker *et al.,* 2012, 2014a, 2014b). In particular, the genomic region containing the single SNP, WUR10000125, was shown to capture 13.2% and 9.1% of the genetic variance in viral load and growth, respectively. Differential expression analyses in the population found associations between the favourable WUR10000125 genotype and a truncated version of the Guanylate Binding Protein (GBP) 5 (Koltes *et al.,* 2015), which seems to play a role in the inhibition of viral replication (Schroyen *et al.,* 2016). Several other associations, with smaller effect sizes were also demonstrated by the PHGC on SSC1, SSC7, SSC17 and SSCX for the Viral Load and Weight gain traits used in these analyses.

Whilst a number of groups have investigated the distribution of PRRS related traits throughout the genome, the results of these findings have been inconsistent with relatively little cross over in the position of the loci identified. Genome-wide association study is a

well-understood and widely used tool for the distribution of trait variance through the genome.

This chapter will investigate the distribution of reproductive effects in these data providing comparison with both the published literature and subsequent regional heritability mapping (see chapter 6). The presence of genetic variance in the reproductive traits having previously been demonstrated in Chapter 3. Any SNPs showing genome wide significance will be characterised further using the measured genotype approach (Boerwinkle *et al.,* 1986), permutation analysis (Churchill & Doerge, 1994), and (where more than one SNP is indicated in a region) linkage disequilibrium (Lewontin & Kojima, 1960).

A wide range of methods have been developed for conducting GWA analyses. The single SNP method which is employed here has several advantages in terms of computational efficiency, reliability, and has been well-quantified in the literature and is flexible in terms of accounting for effects discussed in previous chapters (Iles & Barrett, 2011). As population stratification cannot be ruled out in this study (see results from Chapters 3), two methods which are known to perform well in the presence of stratification were chosen: GRAMMAR (Aulchenko *et al.,* 2007a) and FASTA (Chen & Abecasis, 2007). Both were applied using the genABEL package (Aulchenko, 2011) in R (R Core Team, 2016).

One of the limitations of these and other GWA analysis approaches is the inability to account for repeated records. Whilst a solution to this issue is possible using the GRAMMAR method, for the FASTA methodology, datasets were restricted to only one record per animal in this chapter.

The measured genotype analysis is a widely used validation tool for exploring SNP effects, this method, proposed in Boerwinkle *et al.,* (1986), fits the genotype as a fixed effect in the linear model to estimate both the SNP effect and residual σ^2_A. Adaptations to this model can

additionally explore the SNP interaction to check for independence of the SNP effect of the mixed model terms.

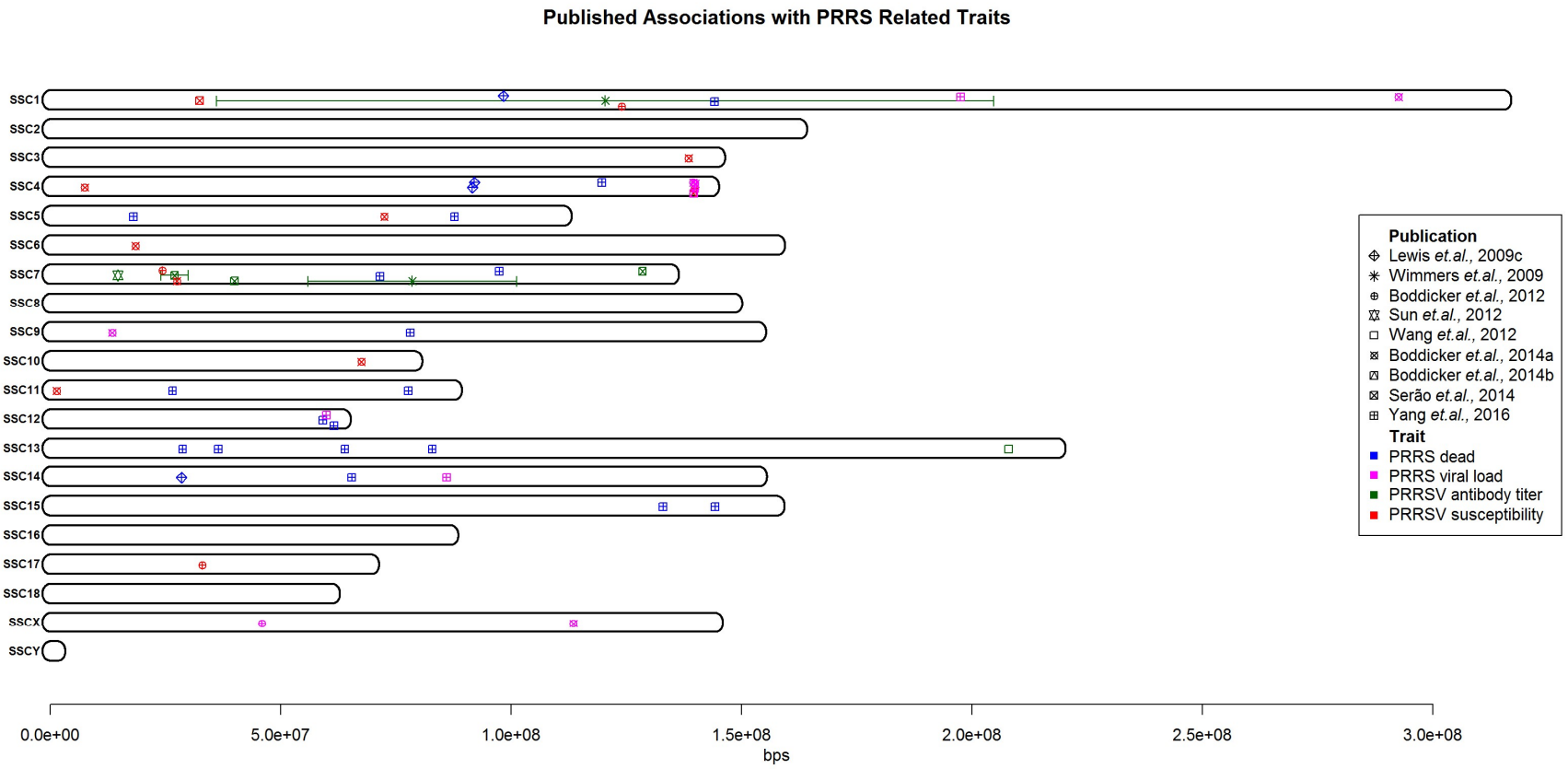

Figure 5.1 – Ideogram Karyotype of the Pig Genome Showing Published Associations for PRRS Traits

Data is collected from the pig QTL database (Hu *et al.,* 2005) with information included from Lewis, (2008); Lewis *et al.,* (2009); Serão *et al.,* (2014) and Yang *et al.,* (2016). Traits are as reported in the database or taken directly from studies. See Text for further details. The indistinguishable points toward at the end of chromosome 4 are a group of papers providing further analysis of the QTL first published for this region in (Boddicker *et al.,* 2012). This plot includes all published associations regardless of significance threshold employed

5.02 Materials and Methods

5.02.1 Data

The datasets are as described previously in Chapters 2 & 3. For summary description of farm 1 data see section for farm 2 see section 2.02.1. For both farms, for a subset of animals, SNP genotypes were available generated using the Illumina PorcineSNP60 chip (Ramos *et al.,* 2009) see also (Illumina, 2015). Mapped locations were used using *Sus scrofa* genome build 10.2, The Swine Genome Sequencing Consortium (Schook *et al.,* 2005).

Data was partitioned into epidemic and non-epidemic phases according to the adapted threshold-threshold method as described in section 2.02.2. The product of the partitioning process was two datasets for Farm 1 containing litters farrowed during PRRSV confirmed Epidemics (Epidemic Phase) and litters farrowed while PRRSV was not detectible by the trend analysis (Non-Epidemic Phase). These two phases were also identified in farm 2, though a further period was identified in the trend analysis showing a period above the threshold not coinciding with a positive herd level PRRSV diagnosis. This period was isolated, and not included in either the non-epidemic or epidemic epidemics representing different examples of epidemic reproductive failure.

Given the inability of GenABEL to account for repeated records, a dataset was created for each phase using only the first record per sow (see section 3.02.1). In farm 2, in keeping with the aims of the research to investigate the genetic effects of PRRSV, where a sow had a litter in the PRRSV confirmed epidemic and the unknown epidemic the litter record within the confirmed epidemic was used.

A joint analysis was also run whereby the data from the two farms was combined by phase (i.e. non-epidemic/epidemic) to look at potential improvements in power as compared to the individual farm analyses.

Quality Control

Quality control was run based solely on the animals (and their corresponding raw SNP genotypes) to be incorporated into each analysis associated with the epidemic / non-epidemic phases. There were slight discrepancies in the recruitment of animals into the analyses according to the method applied. Using GRAMMAR, the environmental residuals used are produced based on the genotypes of all animals in each population (Farm 1, Farm 2 and Joint) that had already passed the quality control procedures (see Chapter 3). For the FASTA and Measured Genotype methods all genotyped animals with phenotype records in each phase are used. This leads to slight discrepancies in the numbers of SNPs, and the number of animals passing quality control between the two methods.

GRAMMAR QC

For the application of the GRAMMAR method, the animals are those used for the polygenic analysis in chapter 3. Given animal have already been subjected to quality control (QC) there were no additional animals excluded in this QC step. Given that; 1) the animals incorporated into the genetic analysis in the previous chapter had already been subjected to QC; and 2) only minor changes are observed in SNP numbers following QC on raw genotypes for these animals: only slight variations are seen in the number of SNPs passing QC for each phase. The results from this quality control procedure are shown for both farms, the joint analysis and all phases in Table 5.1.

Table 5.1 – GRAMMAR Genomic Quality Control

Farm	Raw Nº SNPs	Phase	SNPS Excluded based on		SNPs Passing QC	Number of Animals
			Call Rate (<90%)	MAF (<5%)		
1	60,674	Epidemic	3,529	9,011	48,134	605
		Non-Epidemic	3,529	9,105	48,040	707
2	57,786	Epidemic	355	9,902	47,501	276
			28			
		Epidemic inc. Unknown	395	9,371	47,981	329
			39			
		Non-Epidemic	465	9,557	47,709	585
			55			
Joint	57,440	Epidemic	982	8,335	47,976	881
			147			
		Epidemic inc. Unknown	992	8,172	48,131	934
			145			
		Non-Epidemic	1020	8,295	47,976	1,292
			149			

Results of the quality control run on genotype data for animals with litters for which environmental residuals were generated, by phase. Counts are provided for exclusion of SNPs/animals based on each of the criteria, numbers in merged columns show counts falling into more than one exclusion category (not included in the individual category counts). Exclusion categories show Call Rate, Minor Allele Frequency (MAF). No animal exclusion categories shown as all animals passed QC in this step.

FASTA & Measured Genotype QC

For the FASTA and measured genotype methods identical quality control is run based on the raw number of animals (and their genotypes) which have phenotypic records in each phase (non-epidemic, epidemic and (where applicable) epidemic including unknown), for each population (Farm 1, Farm 2 and Joint). As such a slightly different number of animals and SNPs are used in this analysis. The quality control results for the animals and the quality control results for the SNPs are shown in Table 5.2.

Table 5.2 – FASTA Genomic Quality Control

Farm	Phase	Counts Raw		Counts Falling Under Exclusion Categories					Counts Passing QC	
		SNPs	Animals	SNPs		Animals			SNPS	Animals
				Call Rate (<90%)	MAF (5%)	Call Rate (<90%)	Aut. Het. (FDR<1%)	Corr. IBS (>1)		
1	Epidemic	60,674	640	3,550	9,014	29	0	2	48,087	607
				23		2				
	Non-Epidemic	60,674	748	3,543	9,116	34	0	4	47,994	708
				21		2				
2	Epidemic	57,786	279	381	9,901	3	0	0	47,474	276
				30						
	Epidemic Inc. Unknown	57,786	333	409	9,472	1	3	0	47,866	329
				39						
	Non-Epidemic	57,786	593	476	9,574	3	1	0	47,680	585
				56		4				
Joint	Epidemic	57,440	919	988	8,360	31	0	2	47,943	883
				149		3				
	Epidemic Inc. Unknown	57,440	973	1,001	8,197	32	0	2	48,093	936
				149		3				
	Non-Epidemic	57,440	1,341	1,038	8,309	39	1	4	47,941	1,293
				152		4				

Counts are provided for exclusion of SNPs/animals based on each of the criteria, numbers in merged columns show counts falling into more than one exclusion category (not included in the individual category counts). Exclusion categories show Call Rate, Minor Allele Frequency (MAF), Autosomal Heterozygosity (Aut. Het.) False Discovery Rate (FDR), Corrected Identity By State (Corr. IBS)

5.02.2 Association Study Methodology

The two protocols considered to explore single SNP associations in this study, i.e. FASTA (Chen & Abecasis, 2007) and GRAMMAR (Aulchenko *et al.,* 2007a) both use the genomic relationship matrix to correct for population structure, apply the mixed model methodology in a two-stage process to investigate the significance of SNP associations and are suitable for use in structured populations (Aulchenko, 2011).

Given that individual SNP effects are expected to be small compared to the fixed effects and covariates estimated in the mixed model (Chen & Abecasis, 2007), a polygenic mixed model is fitted in step 1. Step 1 contains no SNP effects, accounting for the various non-genetic effects and obtains the additive genetic variance\covariance matrix and residual error. In step 2 the estimates from step 1 are then used to generate an association test for each SNP. For details of the polygenic model see section 3.02.3. Given that both methods are dependent on the output from the polygenic function; GWA analyses are only considered for traits and models which indicate (P<0.1) a significant heritable effect in the polygenic model in step 1.

FASTA

For the application of FASTA the polygenic mixed model in step 1 is applied using the *polygenic* function in GenABEL which uses maximum likelihood estimation (as compared to restricted maximum likelihood applied in ASreml). This function does not permit the use of multiple records per sow and as such the single record per sow dataset was used. From this function, the genetic variance-covariance matrix (Ω) is obtained as well as a vector of environmental residuals (ê) by animal. From these estimates, a test statistic (T^{Score}) for each SNP (K) is derived thus:

$$T_k^{Score} = \frac{([\boldsymbol{g_k} - E(\boldsymbol{g_k})]' \cdot \boldsymbol{\Omega}^{-1} \cdot \hat{\boldsymbol{e}})^2}{([\boldsymbol{g_k} - E(\boldsymbol{g_k})])' \cdot \boldsymbol{\Omega}^{-1} \cdot [\boldsymbol{g_k} - E(\boldsymbol{g_k})]'}$$

Equation 5.1 – FASTA SNP Association Test Statistic

Where g_k is a vector of the genotypes for SNP k coded 0,1,2 (AA, AB and BB respectively) $E(g_k)$ is a vector of mean-genotypic values (2p where p is the A allele frequency) for SNP k. This derives a test statistic for the genotype whilst removing population structure by modelling the expected correlation between variance components. Both step 1 and step 2 of FASTA were conducted using GenABEL (Aulchenko *et al.,* 2007b) within the R package (R Core Team, 2016).

GRAMMAR

The GRAMMAR approach uses only the vector of environmental residuals (ê) generated using the polygenic mixed model. These are fitted as the response variable in a linear regression on each SNP genotype shown in Equation 5.2:

$$\hat{e} = \mu + g_k + \varepsilon$$

Equation 5.2 – GRAMMAR Environmental Residual Regression on SNP Genotype

Where g_k is a vector of the genotypes for SNP k coded 0,1,2 (AA, AB and BB respectively) and ε is the residual error. Given the dependence on just the environmental residuals from the polygenic model, the output from the ASREML models in Chapter 3 was used. This allows the results from the repeated measures models to be used which indicate a small increase in power to estimate the additive genetic variance as compared to models using the single record data. In this adaptation of the GRAMMAR method used in Lewis et al. (2009) and Bérénos et al. (2015), a mean environmental residual by sow is generated from the environmental residual by litter.

With the polygenic effect removed, these residuals were assumed to be unstructured allowing a basic linear regression on SNP genotype to be performed using the GRAMMAR function in genABEL. Any residual structure will be investigated and, if required, accounted for in subsequent genomic control (GC) steps. Given that the GRAMMAR method can be

liable to underestimate effect size (Aulchenko, 2011), any genome-wide significant SNPs identified using GRAMMAR were validated using a measured a genotype approach described below. With the polygenic effect removed prior to SNP testing, estimation of SNP effects maybe downward biased from GRAMMAR (Aulchenko *et al.,* 2007a).

Measured Genotype

A common method for the validation and further characterisation of SNPs detected using GWA methods is the fitting of an identified SNP as a fixed effect in the polygenic model, also called a measured genotype (MG) analysis (Boerwinkle *et al.,* 1986; Aulchenko, 2011). This enables the significance of an individual SNP genotype on trait variance to be assessed and a prediction of the SNP genotype effect (AA, AB or BB) to be made, whist accounting for the polygenic effect a potential source of bias (Kennedy *et al.,* 1992). A further advantage of the measured genotype method is that confounding between SNP effect and model terms can be investigated, by fitting separate models with interactions between the candidate SNP and potential confounders such as sow line and farm.

The calculation for the SNP variance is given shown in Equation 5.3 (Falconer & Mackay, 1996).

$$\alpha = \frac{\widehat{AA} - \widehat{BB}}{2}$$
$$D = \widehat{AB} - \frac{\widehat{AA} + \widehat{BB}}{2}$$
$$\sigma^2_{SNP} = 2pq\left(\alpha + D(q-p)\right)^2$$

Equation 5.3 – Allele Substitution Effect (ASE), Dominance Effect and SNP Variance Calculations

Where α is the additive/allele substitution effect, D is the dominance and σ^2_{SNP} is the total genetic variance explained by the SNP. $\widehat{AA}$ $\widehat{AB}$ and $\widehat{BB}$ are the predicted values of the major allele homozygote, heterozygote and minor allele homozygote, respectively, p and q are the major and minor allele frequencies, respectively.

Linkage Disequilibrium

Regions identified as showing evidence of association with the reproductive traits under consideration, using the GWA analysis (i.e. more than one SNP indicating a significant effect in a single peak) were explored for linkage disequilibrium (LD) (Lewontin & Kojima, 1960). This, non-random coincidence of alleles at separate loci, was explored using the correlation coefficient (r^2) between the alleles of different SNPs. The r^2 across regions of interest was plotted as a heat density function using the LD heatmap function in R (Shin *et al.,* 2006).

5.02.3 Statistical Analysis

Traits and Models

The traits included in the analysis (*Mum, Still, Dead, Alive, Tof, Fmor, Wean* and *Gest*) are described in section 2.02.1 and summarised in the appendix in Table A.1. Similar to the analyses in chapter 3, for the farm 2 and the joint farm analysis *Wean* was dropped as a trait due to an inability to account for fostering decisions in farm 2 and the traits *Mum, Still and Dead* were transformed using log(*Trait*+1). Given the emphasis on effects attributable to PRRSV infection, disease indicator traits (*Mum, Still, Dead, Alive, Fmor* and *Wean*) are considered. These traits using the defined models, show significant differences (>95 CI) in least square means (LSM) estimates between phases (see Chapter 2) and have low/no heritability estimates in non-epidemic phase with increases seen in the epidemic phase (Chapter 3). The remaining traits (*Tof* and *Gest*), which in many cases indicate a significant heritable effect in both phases are only used to consider the differences in the distribution of SNP effects *between* phases.

The fixed effects and covariates used in the mixed models of this chapter were the same as used in Chapter 3. Three different mixed models (denoted Basic model and alternative models Alt1 and Alt2) were fitted. All three models were used in both the FASTA and

GRAMMAR analysis. Compared to the basic model, alternative model (Alt. 1), fitted additionally a unique identifier for each epidemic, to account for differences between epidemic severities. In the joint analysis given the hierarchical structure Epidemic ID was nested within Farm. To account for the dynamic trend within each epidemic, in alternative model 2 (Alt. 2) an additional trait trend covariate is fitted as an interaction with epidemic ID, additional to the Alt.1 terms. This trend corresponds to the loss trait under consideration (e.g. *Mum, Still* or *Dead*) used in partitioning step. For the other traits (*Alive, Tof, Fmor, Wean* and *Gest*) the *Dead* trend is fitted. Similarly, to Alt.1, in the joint analysis this effect was nested within farm. For analysis including the unknown epidemic only models with the Epidemic ID are considered to ensure differences between, potentially two different sources of reproductive stress, are accounted for.

Power

Consideration was given for the power of the analysis to detect SNP effects in the data. GWAPower is a tool designed to estimate the probability of detecting SNP effects given information on the heritability of the trait, the number of SNPs used, and the number of animals included in the analysis (Feng *et al.,* 2011). Information from the variance component analysis was used with GWAPower to estimate the power of these data to detect SNP effects.

Genomic Control

To consider type I errors the distribution of the test statistic from the genome scan was considered and genomic control was applied (Devlin & Roeder, 1999; Reich & Goldstein, 2001; Zheng *et al.,* 2006). The distribution of the calculated test statistic is compared to a theoretical set of quantiles under a null distribution to look for bulk inflation of the test statistic from the genome scan. These are shown as quantile-quantile plots (Q-Q plots) for each genome scan, where a consistent positive deviation from the line of unity (x=y) provides evidence of inflation. Quantile-quantile plots were also used to investigate other

potential sources of bias and confounding (Zeggini & Morris, 2011). The level of inflation was quantified by the genomic control factor λ (Freedman *et al.,* 2004). Estimation of λ was conducted by regressing the lowest 95% of the observed test statistic onto the theoretical quantiles under a null distribution (Amin *et al.,* 2007). The regression coefficient then provides an estimate of level of inflation in the test statistic. Where $\lambda < 1$ is reported as $\lambda = 1$ (Aulchenko *et al.,* 2007b). To control for the type I error, p values were scaled by the genomic control factor λ, where $\lambda > 1$ (Aulchenko, 2011).

Statistical Inference

The SNP association test statistic from both the GRAMMAR and FASTA methods are assumed to be X^2 distributed with 1 d.f. (Aulchenko *et al.,* 2007a; Chen & Abecasis, 2007).

Bonferroni correction (Dunn, 1959, 1961) was applied to account for multiple testing. An α-value of 0.05 was used throughout the analysis with a corrected significance threshold of $\alpha/nSNPS$ (where nSNPS is the number of SNP effects included in the genome scan). A lower, chromosome level significance was considered as being indicative of an effect, by multiplying the genome-wide significance threshold by 18 (the number of *Sus scrofa* chromosomes.)

Bonferroni correction is a conservative method for correcting for multiple independent tests. This is especially true in GWA methods where the presence of linkage among genotypes can mean that tests are not truly independent.

For all genome scans where any SNPs show a genome-wide significant effect at the Bonferroni corrected threshold; a permutation method was used to estimate empirical, genome wide significance. In this method, the phenotypes are randomly re-assorted among the individuals, breaking down any relationship between genotype and phenotype and creating a test statistic under a null hypothesis of no genetic effect (Churchill & Doerge, 1994). This process is iterated *n* times, and the proportion of null hypothesis test statistics

which exceed those with the genotype/phenotype link maintained provides an empirical P value for each SNP with a precision of 1/*n* (Bush *et al.,* 2012). 10,000 iterations were used in each case. While the permutation method is considered preferable in controlling the type I error rate in GWA studies (Gao *et al.,* 2010), it is computationally demanding, making it impractical across all the genome scans undertaken.

Manhattan plots are shown plotting the $-\log_{10}$(p-values) from the genome scans under the two methods.

In the measured genotype analysis, the effect of the SNP genotype on trait variance was considered using the Wald F statistic. Using the additive substitution effect or dominance effect, in conjunction with the corresponding standard error of the difference (SED), a t test was constructed to test the significance of the effect size under each model.

5.03 Results

5.03.1 Farm 1

Using the heritability estimates from the variance component analysis with ~48,100 SNPs and ~606 individuals, the power using the basic model was estimated as 1 (*Still* and *Wean* trait) and 0.93 (for the *Mum* trait, the lowest basic model heritability estimate), suggesting adequate power to detect SNP effects. When alternative model 1 was considered this reduced to 0.81 for *Mum* and *Alive* traits (the lowest heritability estimates indicating significance), and 1.00 for *Wean*, suggesting sufficient power to detect SNP effects for all those traits for which a significant heritable component could be demonstrated. No significant heritable component could be estimated using Alt.2 model.

FASTA Results

The FASTA method was applied to the single record data for all reproductive traits. In step one, estimation of the polygenic effect in GenABEL failed to identify a heritable component

for those traits which show low component to standard error ratios (<2) estimated in Chapter 3. This is most likely the result of a reduction of the number of records in the single record data, relatively high standard errors, and a bias toward the fixed effects and covariates inherent to the maximum likelihood estimation implemented in genABEL (as compared to the restricted maximum likelihood embedded in ASREML).

Table 5.3 shows the resulting heritability estimates for the *Still, Tof, Wean* and *Gest* traits. Two disease indicator traits (*Still* and *Wean*) both show heritable effects in the epidemic phase, traits not shown failed to converge or provided estimates of heritability $<1\times10^{-8}$ (P=1). Using the *Gest* trait a significant heritable component could be estimated in both phases allowing for direct comparison of the distribution of effects in each phase. A heritable effect is only observed in the non-epidemic phase for *Tof.*

Table 5.3 – Heritability Estimates and Associated P Values Generated Using Polygenic Function in genABEL.

Trait	Model	Heritability h^2 and (P value) Calculated by Phase	
		Epidemic	Non-Epidemic
Still	Basic	0.09 (0.06)	–
	Alt.1	–	NA
	Alt.2	–	–
Tof	Basic	–	0.12 (3×10^{-3})
	Alt.1	–	NA
	Alt.2	–	0.12 (2×10^{-3})
Gest	Basic	0.18 (2×10^{-3})	0.17 (8×10^{-5})
	Alt.1	0.13 (0.05)	NA
	Alt.2	0.14 (0.03)	0.16 (2×10^{-4})
Wean	Basic	0.28 (2×10^{-8})	–
	Alt.1	–	NA
	Alt.2	–	–

Heritability and (P-values) estimated using the *polygenic* function in genABEL as applied in step 1 of the FASTA methodology. "NA" indicates model not applicable to that phase, "–" indicates estimate $<1\times10^{-8}$ (P=1) or failed to converge

Based on the variance component estimates from these models, SNP associations were tested using the FASTA method using the *mmscore* function in genABEL. Epidemic phase, basic model Manhattan plots, including corresponding QQ plots (showing estimate of λ) are shown for the *Still* and *Wean* in Figure 5.2.

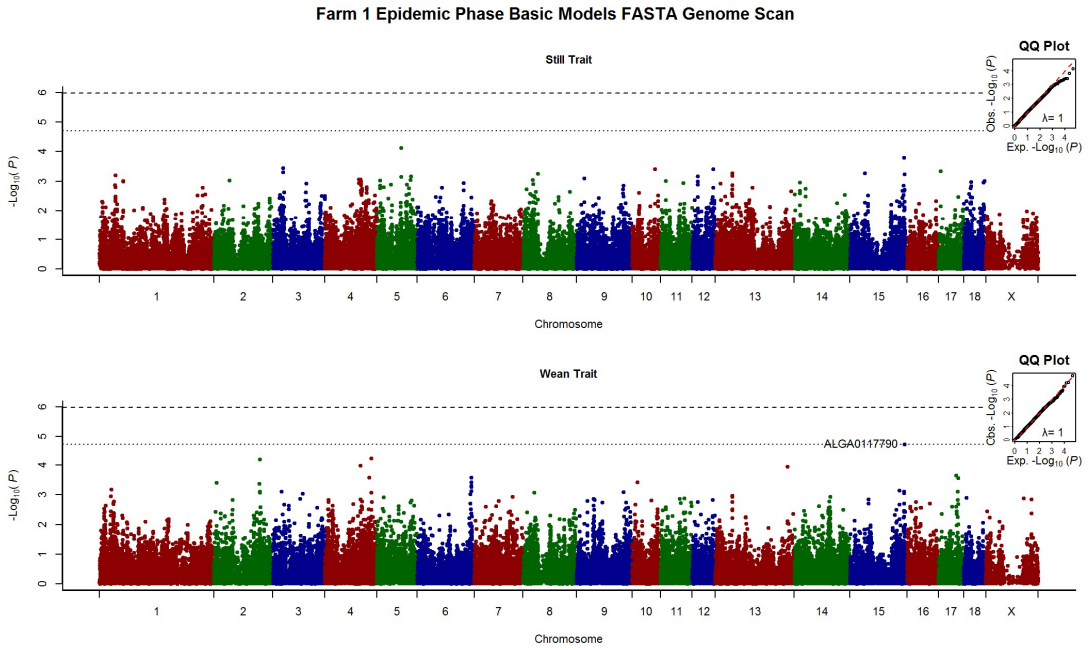

Figure 5.2 – Farm 1 Manhattan Plots using FASTA for Still and Wean Traits Using the Basic Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,087 SNPs. 1 SNP significant at the chromosome level significance threshold shown labelled. QQ plots show x=y (red dashed.

As seen in the QQ plots in the above figure, inflation and confounding did not appear to be present, λ reported at 1, and no regular aberrations in the distribution of the test statistic are observed. Only one SNP indicates an effect at the reduced (chromosome level) significance threshold for the *Wean* trait; ALGA0117790, located at the end of SSC 15 (15:150247212). Interestingly a signal in this region is alluded to in the *Still* trait analysis shown by a trail of SNPs at a similar approximate location, though a different top SNP in this region is identified as ALGA0119312 (15:149350761), approximately 8.96 kilobase pairs (kbp) from the SNP indicated in the *Wean* trait analysis.

In contrast to disease indicator traits, the *Gest* trait shows a moderate heritable genetic effect in both epidemic and non-epidemic phase; allowing for a comparison of the distribution of SNP effects by phase (Figure 3).

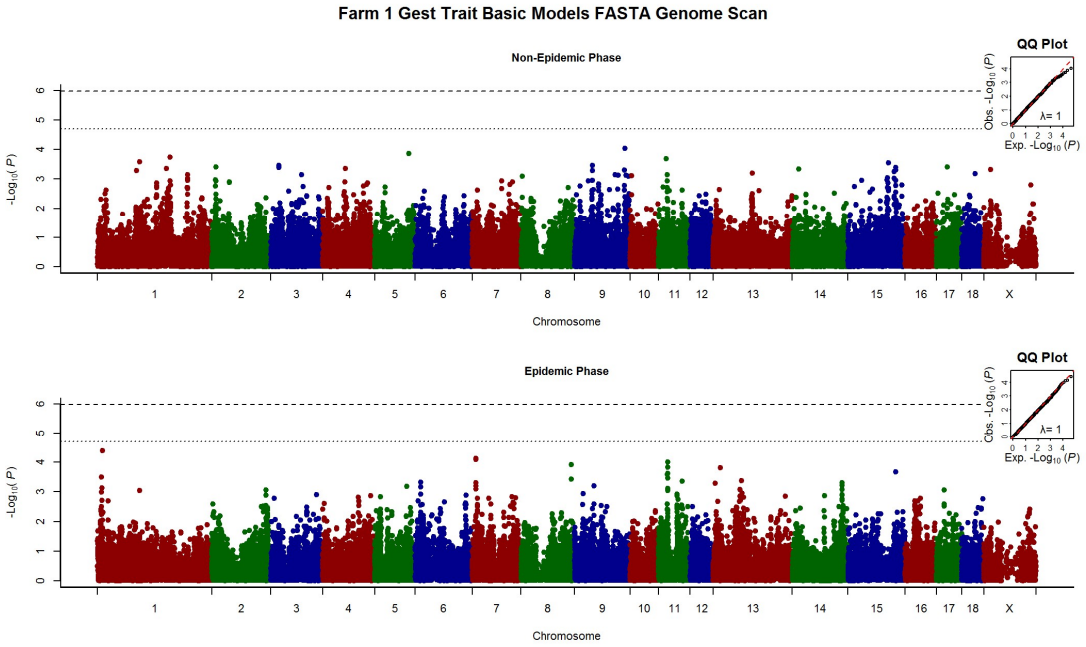

Figure 5.3 – Farm 1 Manhattan Plots using FASTA for Gest Trait Using the Basic Model with Data from Each Phase

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,087 SNPs. QQ plots show x=y (red dashed line).

No SNPs reached genome-wide or chromosome level significance. The highest peak on SSC11 is at ~24Mbp, common to both phases though most peaks are distinct between the two plots suggesting that the two phases are regulated by different sets of SNPs.

While one SNP (ALGA0119312) indicates an effect in the *Wean* trait basic model FASTA analysis on SSC15, only a slight indication is seen for the presence of a SNP effect in the QQ plot as a deviation above the line of unity for the indicated SNP. Whilst there is limited evidence of an effect in both the *Still* and *Wean* traits, given a lack of strong evidence within an individual genome scan these were not characterised further as part of this analysis.

GRAMMAR Results

All reproductive performance traits in the Epidemic phase, under the basic model indicated a significant heritable component; in Non-Epidemic phase only the *Mum*, *Tof*, *Wean* and *Gest* traits indicate a significant heritable component (see Table 3.3). The resulting $-\log_{10}$ GRAMMAR SNP P values, using the basic model, for all disease indicator traits are shown in Figure 5.4.

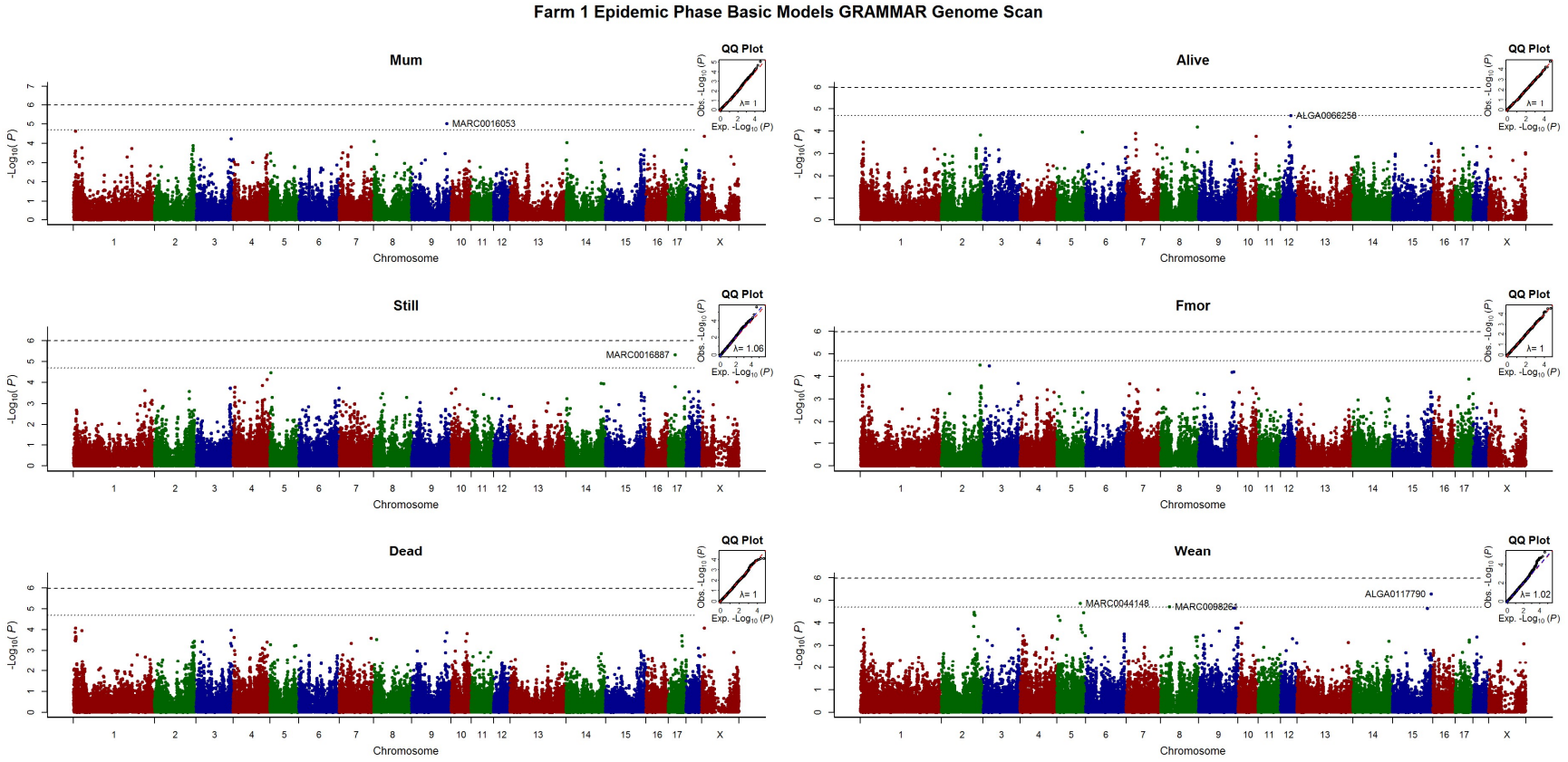

Figure 5.4 – Farm 1 Manhattan Plots using GRAMMAR for Disease Indicator Traits Using the Basic Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,134 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

Of these six analyses, no SNPs are significant at the genome-wide level, six SNPs show potential effects indicated at the chromosome level of significance on SSC9 (*Mum*), SSC12 (*Alive*) and SSC17 (*Still*) with three regions indicated for the *Wean* trait on SSC5, SSC8 and SSC1. Whilst limited amounts of inflation are indicated in the *Still* and *Wean* traits, it is within acceptable levels and controlled for applying genomic control where $\lambda>1$. In all the QQ plots which have indicative associations in the corresponding Manhattan plots, true SNP effects are indicated, even when GC is accounted for (blue dashed line), by the deviation from the null distribution in a small number of SNPs in the upper end of the distribution.

A significant effect on SSC9 at the chromosome level significance threshold for the *Mum* trait is also indicated for the *Wean* trait, albeit for different SNPs (MARC0016053 and ASGA0044703 respectively, with a distance of 1.9Mbs between the two). The effects on SSC12 for the *Alive* trait, and on SSC17 indicated for the *Still* trait, are only indicated at the chromosome threshold for each trait individually. The effect on SSC15 is at the same location as identified by the FASTA method: both methods indicated ALGA0117790 as significant SNP for the *Wean* trait.

In the non-epidemic phase disease two indicator traits (*Mum* and *Wean*) were demonstrated to have significant heritable effects. The respective GRAMMAR scans (Appendix Figure A.7) show peaks at very different location to those shown in the epidemic phase analysis. The *Mum* trait genome scan was affected by high levels of inflation in the test statistic ($\lambda=1.3$) which could indicate confounded variables and/or residual structure in the environmental residuals. No additional information with which to explore structures not accounted for in the mixed model is available. As such whilst a single genome-wide significant SNP (ASGA0036173) is observed on SSC7 for the *Mum* trait this was not pursued.

Using the alternative model 1 only the *Still*, *Fmor* and *Wean* traits show a significant heritable effect in the variance component analysis. These GRAMMAR genome scans are shown in Figure 5.5.

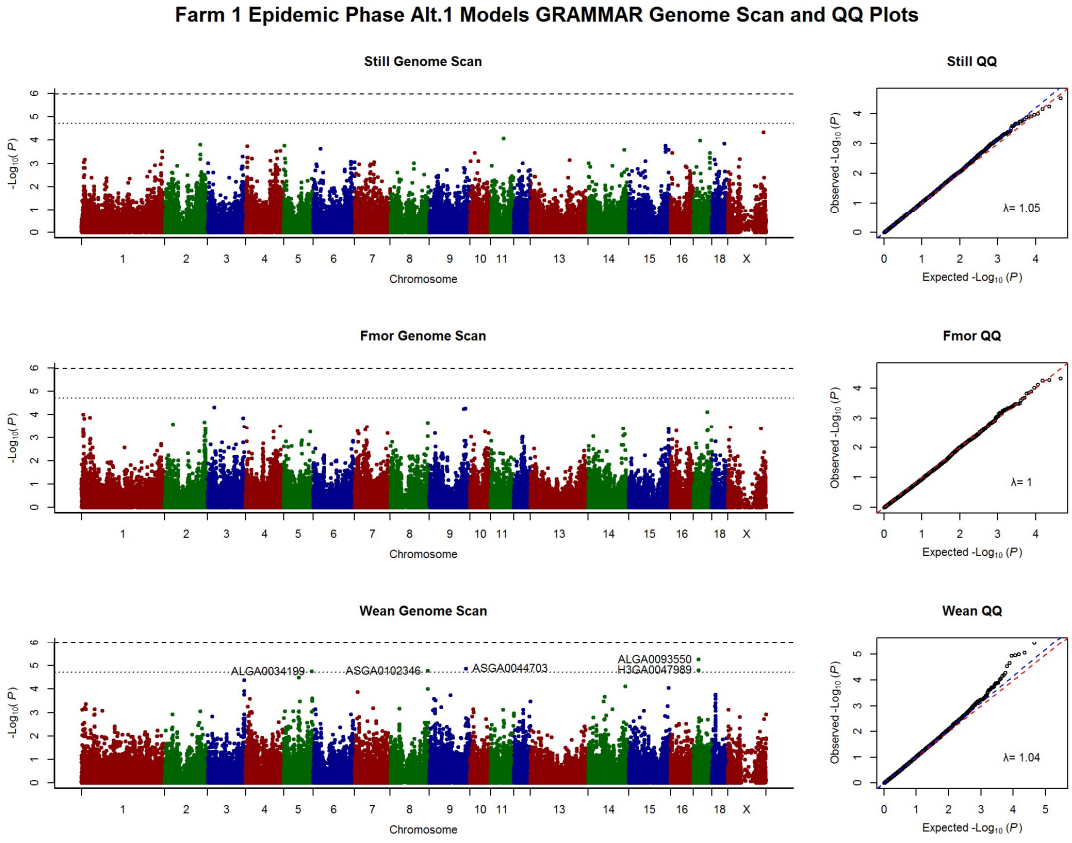

Figure 5.5 – Farm 1 Manhattan Plots using GRAMMAR for Still, Fmor and Wean Traits Using the Alt.1 Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,134 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

No significant SNP effects are indicated in the *Still* and *Fmor* traits in the genome scan above either significance threshold, or in the QQ plots by a deviation from the null distribution at the upper extreme (Figure 5). Inflation increases slightly for the *Wean* trait as compared to the basic model (λ=1.04 cf. 1.02) though is still within acceptable levels. Inflation is accounted for in the reported -$\log_{10}$(P); the extent of this scaling is shown in the QQ plot (blue dashed line) as compared to the line of unity (red dashed line). Four regions (5

SNPs) are indicated in the *Wean* trait genome scan. Two SNPs are significant above the chromosome level threshold on SSC15. ALGA0093550 and H3GA0047989, these are adjacent SNPs in the analysis with a high r^2 value of 0.96. ALGA0034199 shown at the chromosome level of significance on SSC 5 using the Alt.1 model is ~16Mbp downstream from the SNP indicated using the basic model, however using the basic model a SNP just below the chromosome level significance threshold (ALGA0033856) is only 4Mbp away from ALGA0034199. The SNP significant on SSC9 using the Alt.1 model (ASGA0044703) appears just below the chromosome level of significance for the basic model (4^{th} SNP by rank). To consider whether effect for other traits were equally sensitive to changes between the models the top 5 SNPs by rank for basic model were considered.For the Still trait of the 5 top SNPs by rank in the basic model 2 of them appear in the top 5 SNPs by rank in alternative model two, for the *Fmor* trait this increased to 4. This suggests, whilst a degree of sensitivity is seen to alternative model terms in the location of SNPs there is considerable consistency across models.

No SNPs show significance at the genome wide level, as such a significant effect at a specific locus cannot be demonstrated in these data. Limited amounts of evidence are seen suggesting indicating regions of interest. A SNP on SSC15 shows chromosome level significance in both the FASTA and GRAMMAR methods. Additionally, two SNPs in high LD, show an association with the Wean trait on SSC17, using the alternative 1 model these are not seen in the basic model. Whilst some indication of consistent signals between Basic and Alt.1 models (such as is seen for the *Wean* trait at the end of SSC5) the distribution of SNP effects at some loci appears sensitive to the inclusion of epidemic ID. Given the differences seen in disease indicator traits between phases shown in chapters 2 and 3 considerable differences were expected between phases in the genome scans as shown for the *Wean* and *Mum* traits.

5.03.2 Farm 2

Using the information from the variance component estimates with ~47,794 SNPs and ~276 individuals, power was estimated using the basic model at 1.00 for the *Still* trait, the only trait for which a significant heritable effect is indicated, suggesting adequate power to detect SNP effects. When alternative model 2 was considered this remained the same for the *Still, Alive* and *Fmor* traits (showing a significant heritable component) again suggesting sufficient power to detect SNP effects.

FASTA

The polygenic effect was calculated in step one. The resulting heritability estimates and p values for the polygenic effect are shown in Table 5.4.

Table 5.4 – Heritability Estimates and Associated P Values Generated Using Polygenic Function in genABEL.

Trait	Model	Heritability h^2 and (P) Calculated by Phase		
		Epidemic	Epidemic Including Unknown	Non-Epidemic
Mum	Basic	0.07 (0.29)	0.05 (0.38)	0.04 (0.27)
	Alt.1	NA	–	NA
	Alt.2	0.12 (0.18)	0.05 (0.61)	0.04 (0.25)
Still	Basic	0.13 (0.07)	0.15 (0.02)	0.04 (0.05)
	Alt.1	NA	0.15 (0.02)	NA
	Alt.2	0.14 (0.07)	0.15 (0.03)	0.04 (0.05)
Dead	Basic	–	–	0.05 (0.06)
	Alt.1	NA	–	NA
	Alt.2	–	–	0.05 (0.05)
Alive	Basic	0.08 (0.19)	0.13 (0.03)	0.01 (0.36)
	Alt.1	NA	0.13 (0.03)	NA
	Alt.2	0.13 (0.06)	0.16 (0.01)	0.02 (0.27)
Tof	Basic	–	0.04 (0.29)	0.05 (0.11)
	Alt.1	NA	0.03 (0.42)	NA
	Alt.2	–	0.03 (0.4)	0.05 (0.08)
Fmor	Basic	0.04 (0.41)	0.1 (0.08)	0.04 (0.11)
	Alt.1	NA	0.1 (0.07)	NA
	Alt.2	0.12 (0.11)	0.15 (0.02)	0.04 (0.1)
Gest	Basic	–	–	0.23 (5×10^{-4})
	Alt.1	NA	–	NA
	Alt.2	–	–	0.21 (1×10^{-3})

Heritability and (P-values) estimated using the *polygenic* function in genABEL as applied in step 1 of the FASTA methodology. "NA" indicates model not applicable to that phase, "–" indicates h^2 estimate $<1\times10^{-8}$ (P=1) or failed to converge.

Similar to the results of ASREML analysis, low numbers of animals are thought to limit the ability demonstrate a significant heritable effect for some traits and models. Where a polygenic effect is both estimable and indicated (P<0.1) a similar pattern is observed in these

estimates as is seen in chapter 3, in that an increase in heritability is observed in the epidemic phases as compared to non-epidemic phase. In many cases a larger heritability estimate is generated under alternative model 2 than under the basic model. For the *Fmor* trait a significant (P<0.05) heritable component is only indicated under alternative model 2.

Applying the FASTA methodology to those traits and models for which a significant heritable component could be demonstrated, failed to identify any SNPs significantly associated with the reproductive traits in most cases. A single SNP indicated an effect for the *Still* trait during the epidemic including unknown phase under models Alt.1 and Alt.2. The Manhattan plot for the FASTA genome scan of the *Still* trait using the Alt.1 and Alt.2 model is shown in Figure 5.6. The basic model was not considered for epidemic including unknown phase given the unknown source of reproductive stress in the unknown epidemic. The Manhattan and QQ plots for the *Alive* trait analysis which failed to suggest any association are shown in the appendix.

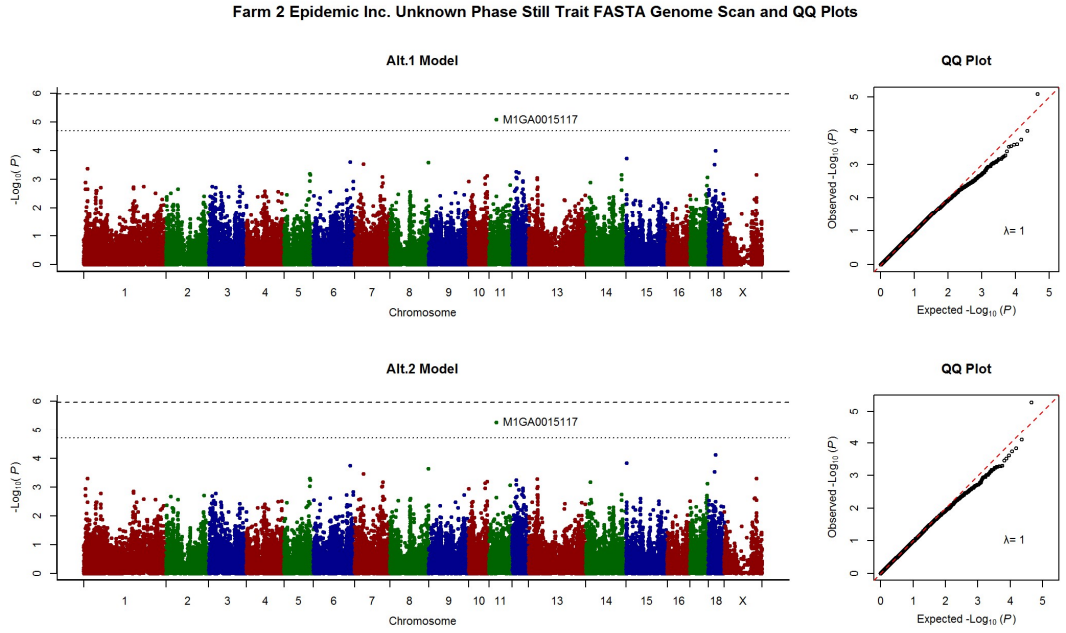

Figure 5.6 – Farm 2 Manhattan Plots using FASTA for Still Trait Using the Alt.1 and Alt.2 Models with Epidemic Including Unknown Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,866 SNPs. SNPs significant above the

indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

While the potential for a true effect is indicated in the QQ plot (most SNPs showing no effect, with the one top SNP showing a deviation from null distribution), given a lack of supporting evidence, such as a trail of increased $-\log_{10}$(P values) in the region, this effect was not explored further.

As can be seen from the plots a very similar distribution of the SNP effects is seen between the Alt.1 and Alt.2 models.

GRAMMAR Results

Following GRAMMAR genome scans on the environmental residuals produced using the Epidemic phase data resulted in high λ estimates. The QQ plots for traits and models for which a heritable component was indicated (P<0.1) in the ASREML analysis (see Table 3.7) are shown in Figure 5.7.

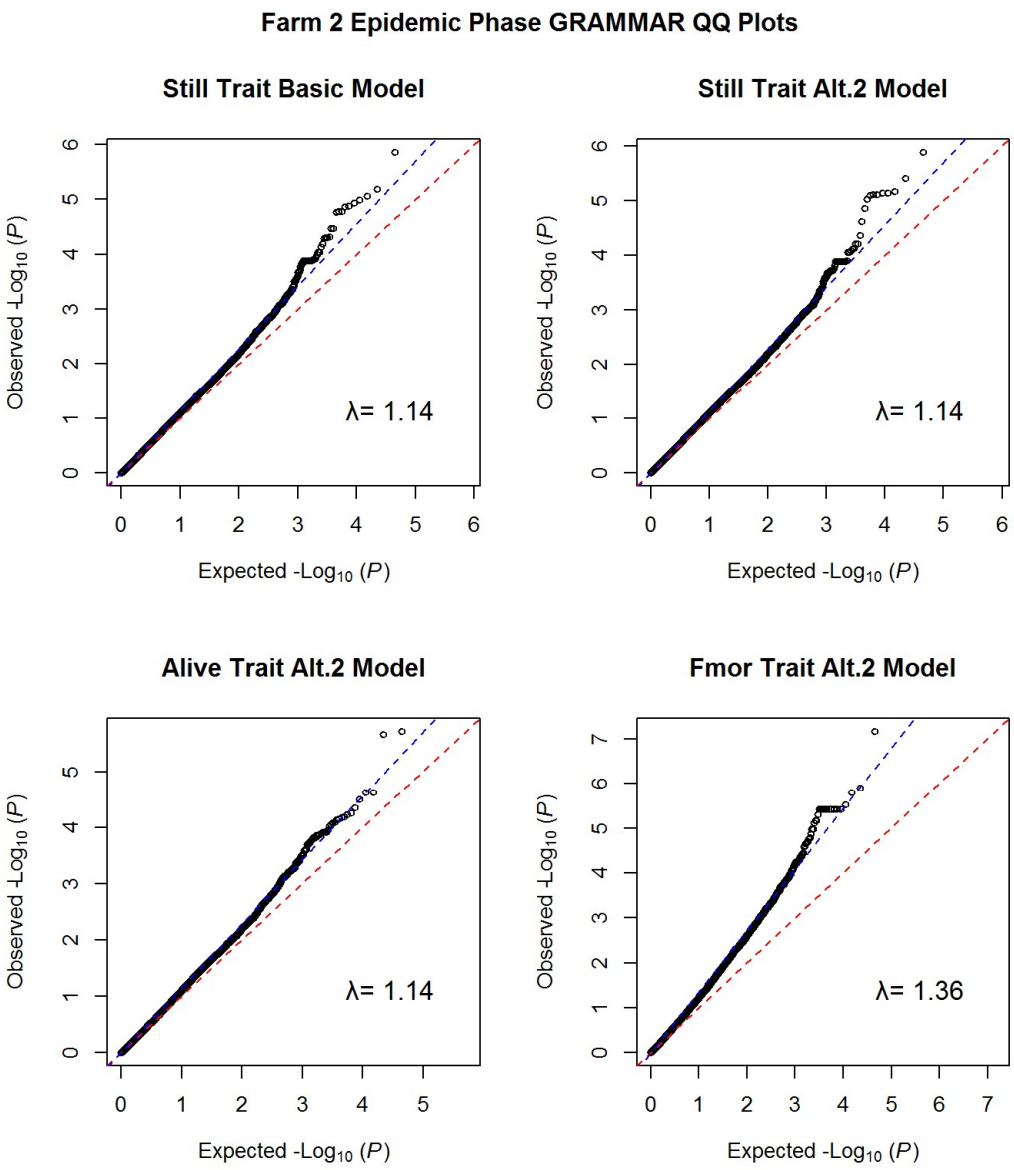

Figure 5.7 – Farm 2 Epidemic Phase GRAMMAR QQ Plots

Expected and observed -$\log_{10}$(P) values from the GRAMMAR score applied to the Farm 2 Epidemic phase analysis. Traits and models shown indicate (P<0.1) a heritable component in the ASREML analysis in Table 3.7. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

In Figure 5.7 inflation of the test statistic is indicated in the QQ plots and estimates of lambda. High levels of inflation can be an indication of confounding in the statistical models. This confounding can occur from a variety of sources including incomplete accounting of environmental sources of variation, population stratification or genetic structures created by selection. Very high p-values in specific regions could indicate regions targeted by selective

breeding and resulting loss in diversity (Campbell *et al.,* 2005) though very few SNPs met chromosome level significance and inflation did not appear restricted to specific regions. To consider the effect of population structure on inflation, principle components were considered and included in the ASREML model (see section 3.02.3). The inclusion of principal components slightly increased the λ estimates suggesting population structure was not responsible for the observed inflation. It is possible that residual structure in the environmental residuals occurred as a result of an incomplete accounting for the polygenic effect, given the low numbers of animals involved and high standard errors for the additive genetic variance.

As shown in Figure 5.8 when the additional data from the unknown epidemic is included, a slight reduction is seen in the estimates of lambda. An increase is also seen in the component to standard error ratio when the repeated measures model is used (see Table 3.7) as compared to the single measures data (see Table A.7) suggesting improvements in statistical power. Whilst lambda is still elevated this is accounted for in the scaling of the SNP p-values using genomic control.

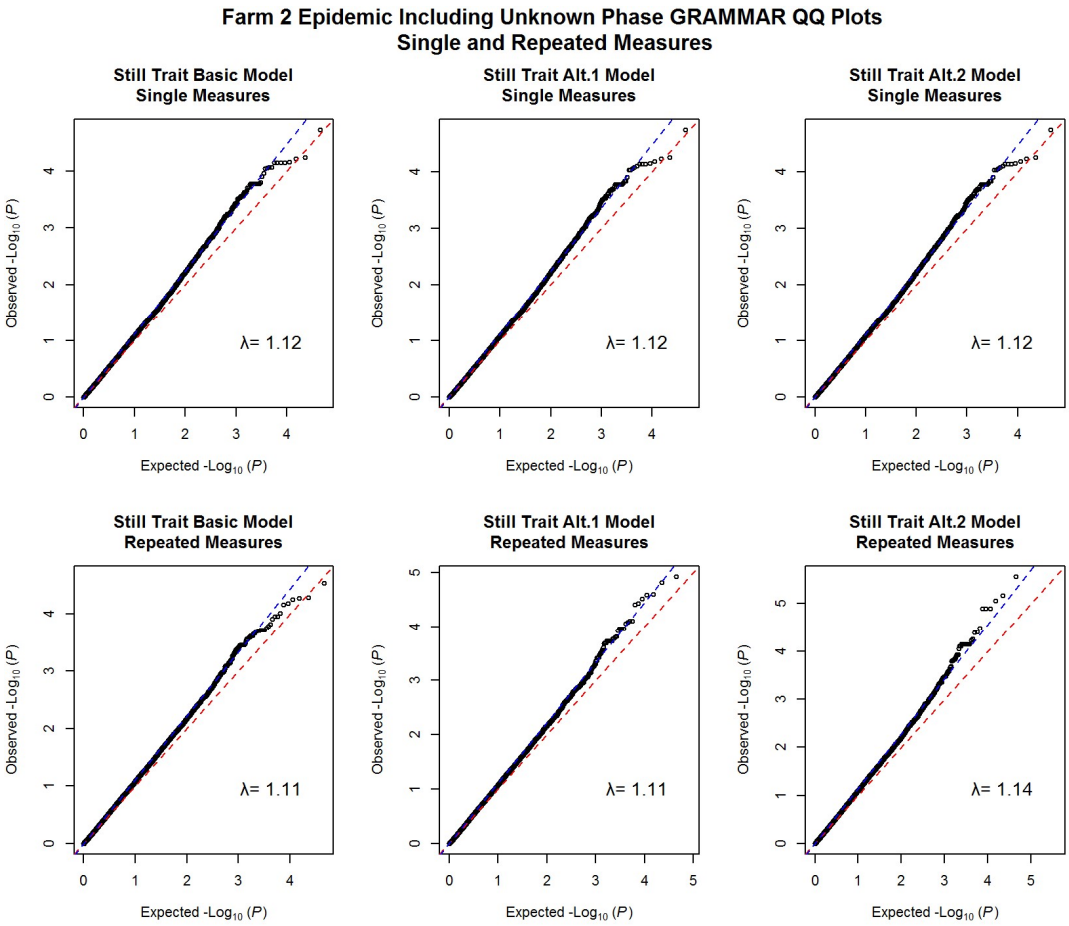

Figure 5.8 – Farm 2 Epidemic Including Unknown Phase GRAMMAR QQ Plots

Observed -$\log_{10}$(P) values from the GRAMMAR analysis compared to a theoretical set of quantiles under a null distribution for repeated and single measures dead. x=y (red dashed line) where $\lambda>1$ degree of inflation indicated (blue dashed line)

The Manhattan plots of the GRAMMAR genome scan using *Still* trait environmental residuals across phases and models are shown in Figure 5.9. For the Epidemic and Non-Epidemic phases the basic model is shown, for the epidemic including unknown phase alternative model 1 is shown; alongside alternative model 2 genome scans for all phases. This shows the contrast including (left) and omitting (right) the *Still* trait trend.

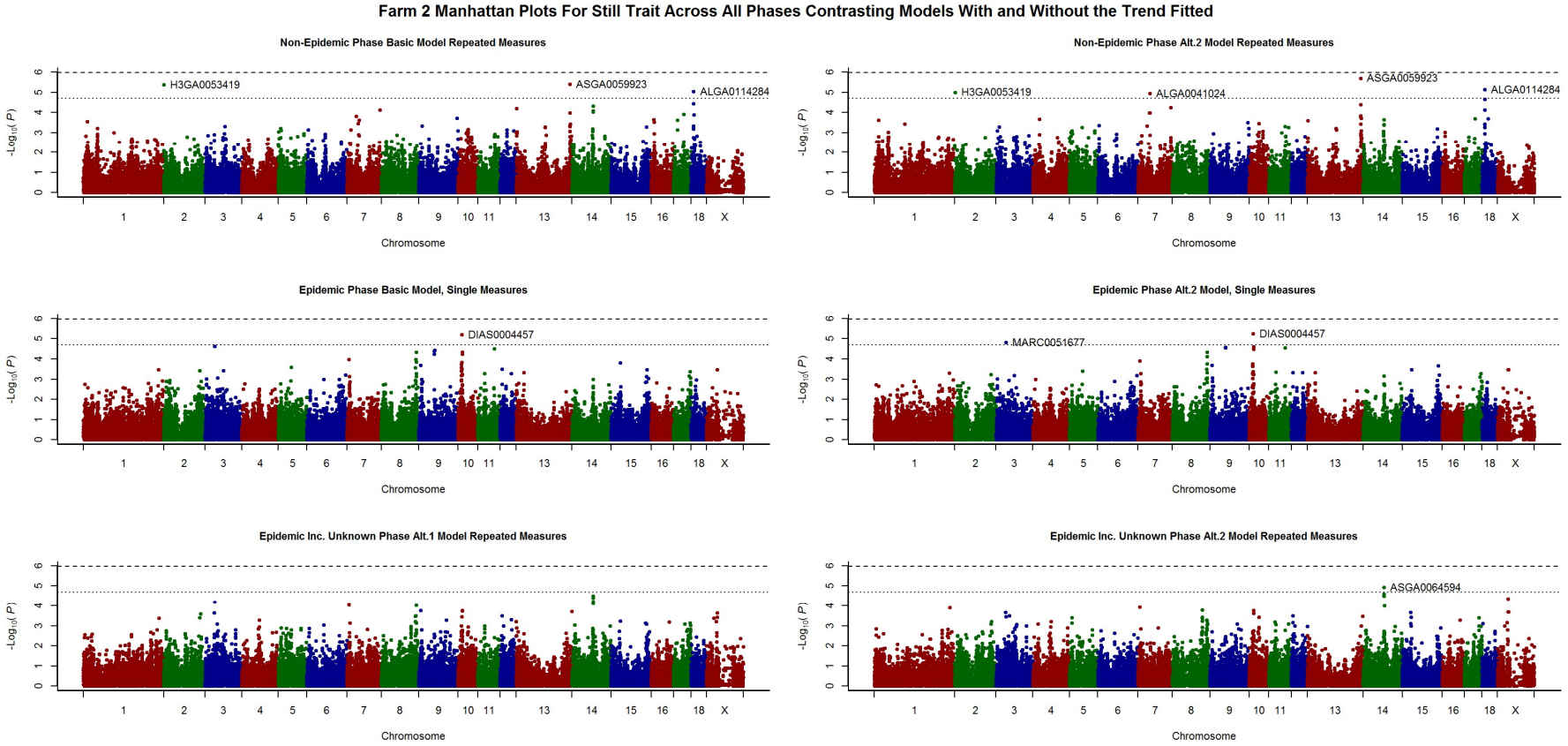

Figure 5.9 – Farm 2 Manhattan Plots using FASTA for Still Trait Using the Models Including and Omitting Trait Trend All Phases Shown

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line) based on 47,709, 47,501 or 47,981 SNPs for epidemic, epidemic inc unknown and non-epidemic respectively. SNPs. SNPs significant above the indicative level shown labelled. QQ plots shown for non-epidemic (appendix Figure A.8), epidemic phase (Figure 5.7), epidemic inc. unknown phase (Figure 5.8). QQ plots show x=y (red dashed line) where $\lambda>1$ degree of inflation indicated (blue dashed line)

A marked difference is seen in the distribution of effects between the non-epidemic and epidemic phases, though a common feature to all scans is evidence of an effect on SSC14, which reaches chromosome wide significance for epidemic including unknown phase in the Alt 2 model (SNP ASGA0054954) (Figure 10). In these data no difference is seen fitting the trait trend in the general distribution of the effects other than minor differences in the reported P values.

The peaks on SSC13 (ASGA0059923) and SSC18 (ALGA0114284), which are significant at the chromosome level threshold in the non-epidemic phase are not indicated in epidemic phase analysis, while the peak on SSC10 (DIAS0004457), which is significant at the chromosome level in the epidemic phase analyses is not indicated in the non-epidemic phase analyses. It is interesting that with the inclusion of 54 litters from the time-period of increased reproductive stress the top SNP changes from DIAS004457 on SSC10 to ASGA0064594 on SSC14.

5.03.3 Joint Analysis

Using GWAPower, power was estimated at > 0.98 for all traits in the epidemic phase under the basic model and alternative model Alt1.

FASTA Results

A significant ($p<0.05$) heritable component was only observed for some reproductive traits, namely *Still*, *Tof* and *Gest,* for some models and phases. These are shown in Table 5.5.

Table 5.5 – Heritability Estimates and Associated p-Values Generated Using The Polygenic Function in genABEL.

Trait	Mode	Heritability h^2 and (P) Calculated by Phase		
		Epidemic	Epidemic Inc. Unknown	Non-Epidemic
Still	Basic	0.13 (0.04)	0.13 (0.03)	–
	Alt.1	0.09 (0.38)	0.09 (0.27)	NA
	Alt.2	–	–	–
Tof	Basic	0.08 (0.01)	0.07 (0.02)	–
	Alt.1	0.06 (0.04)	0.06 (0.04)	NA
	Alt.2	0.06 (0.04)	0.06 (0.04)	0.04 (0.24)
Gest	Basic	0.16 (4×10^{-3})	0.13 (0.02)	0.18 (1×10^{-8})
	Alt.1	0.13 (0.03)	0.10 (0.14)	NA
	Alt.2	0.12 (0.05)	0.10 (0.17)	0.18 (2×10^{-8})

Heritability and (P-values) estimated using the polygenic function in genABEL as applied in step 1 of the FASTA methodology.

The subsequent FASTA genome scan indicated a single SNP reaching chromosome wide significance for the *Still* trait in the Epidemic phase under the basic model, with very similar Manhattan plots for the alternative model Alt.1 and Alt.1 model when the unknown epidemic was included (Figure 5.10). This SNP (ASGA0105527) is the top SNP in all genome scans shown)

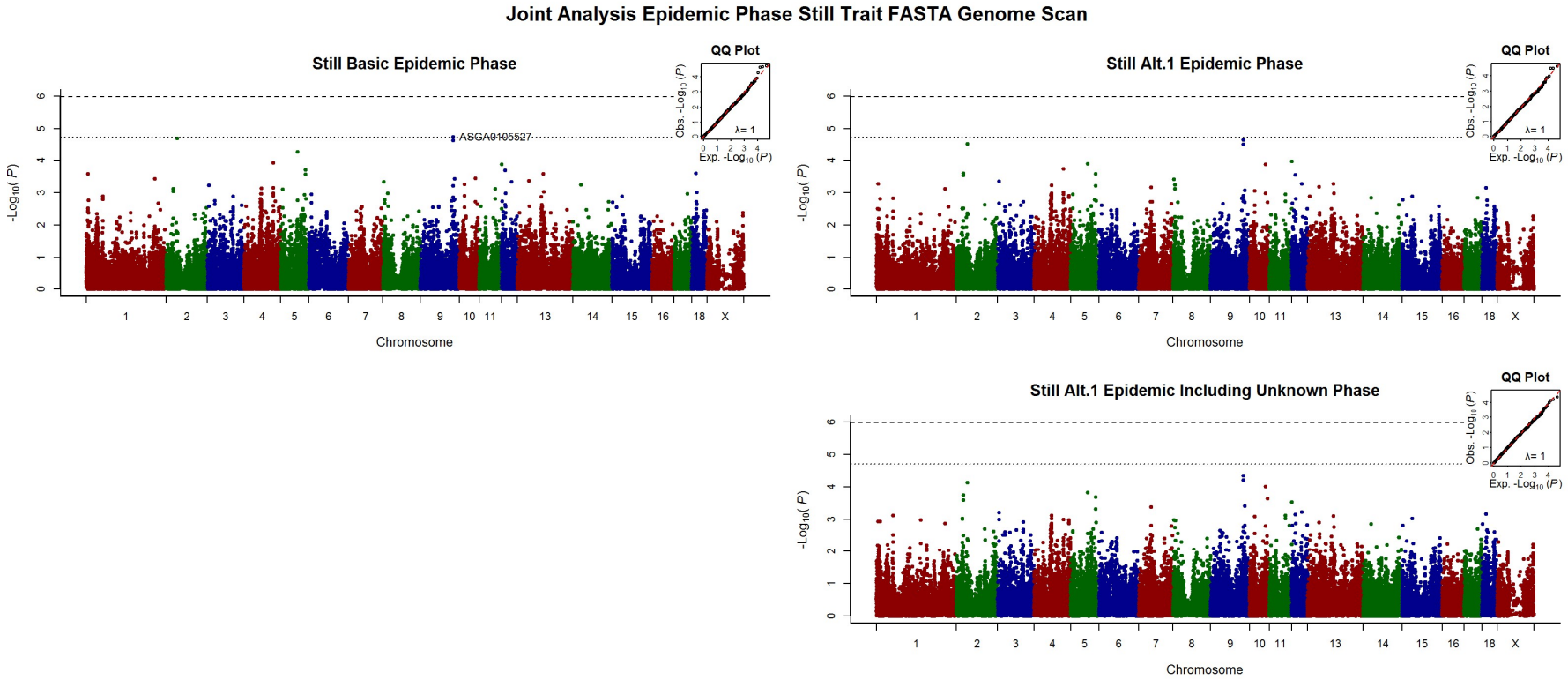

Figure 5.10 – Joint Farm Manhattan Plots using FASTA for Still Traits with Epidemic and Epidemic Including Unknown Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,943 for epidemic phase and 48,093 for epidemic including unknown phase. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

GRAMMAR Results

The variance component estimates from the mixed models used to derive the environmental residuals for the joint farm GRAMMAR analysis are shown for the non-epidemic and epidemic phase in Table 3.8 and for the epidemic including unknown phase in Table 3.9.

Epidemic phase

For the epidemic phase most reproductive performance traits (*Mum, Still, Dead, Alive,* and *Fmor*) show a significant ($P<0.05$) or close to significant heritable component under the basic model. Figure 5.11 shows the $-\log_{10}(P)$ of SNPs following GRAMMAR regression of the joint farm genotypes on the environmental residuals from the basic model analysis.

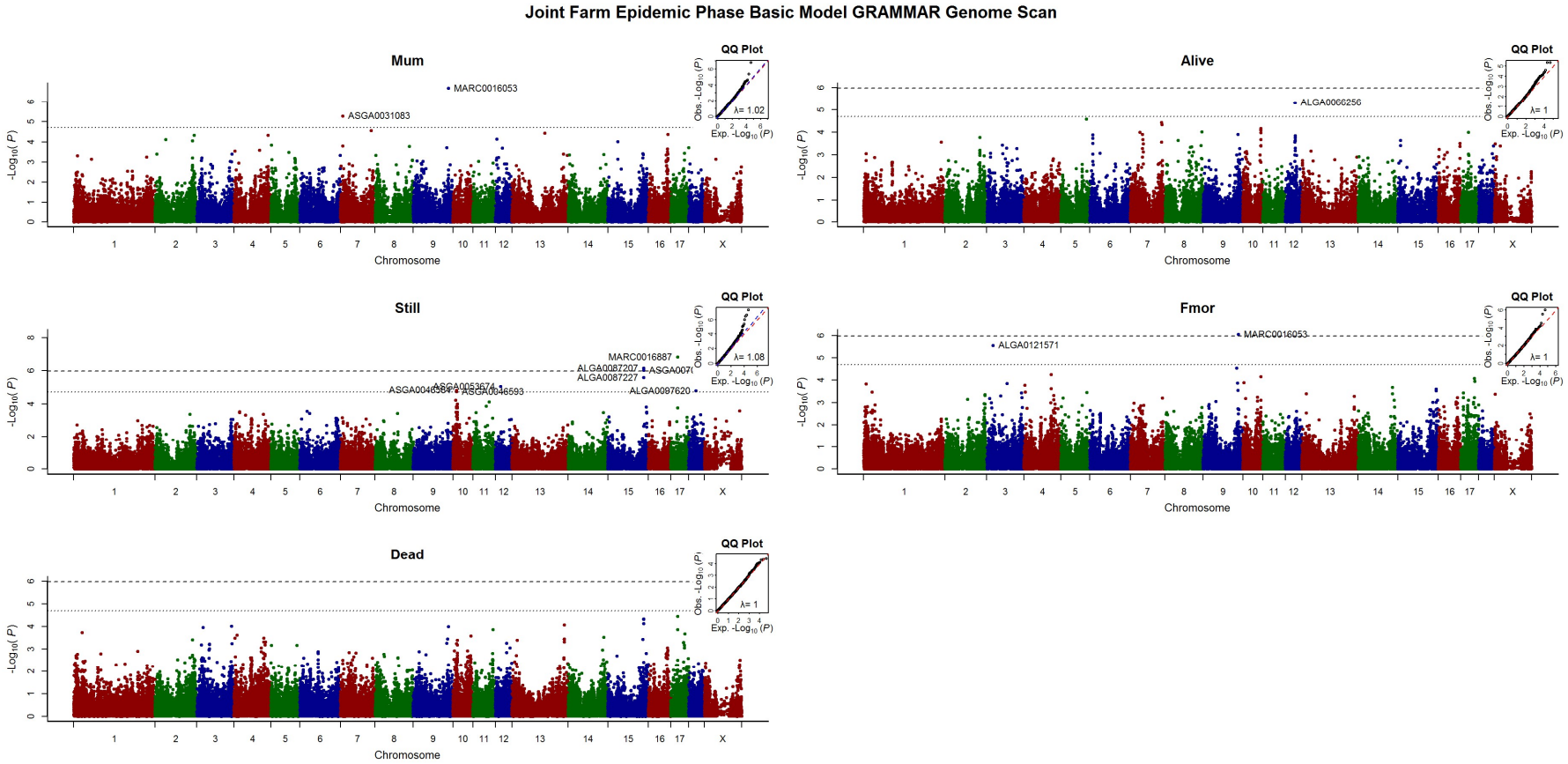

Figure 5.11 – Joint Farm Manhattan Plots using GRAMMAR for All Disease Indicator Traits Using the Basic Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,976 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

Across the traits 13 SNPS show evidence of a significant effect over the chromosome level significance threshold, of which 5 show an effect at the genome wide corrected threshold. One SNP (MARC0016053) appears in both the *Mum* and *Fmor* analyses. *Alive*, *Fmor* and *Dead* show no evidence of inflation (λ=1) while *Still* and *Mum* show limited inflation in the test statistic at λ=1.02 and λ = 1.08, which is accounted for by the scaling of the reported P-values. In all genome scans except for the *Dead* trait, a positive deviation from the null distribution is observed at the upper end of the QQ plots as expected for a limited number of SNPs showing effects.

For the Alt.1 model in the variance component analysis a slight reduction in power to estimate the heritable genetic component is seen compared to the basic model. However, a genetic effect is indicated (P<0.1) in all traits and the results from the GRAMMAR single SNP analysis of the environmental residuals from this model is shown in Figure 5.12.

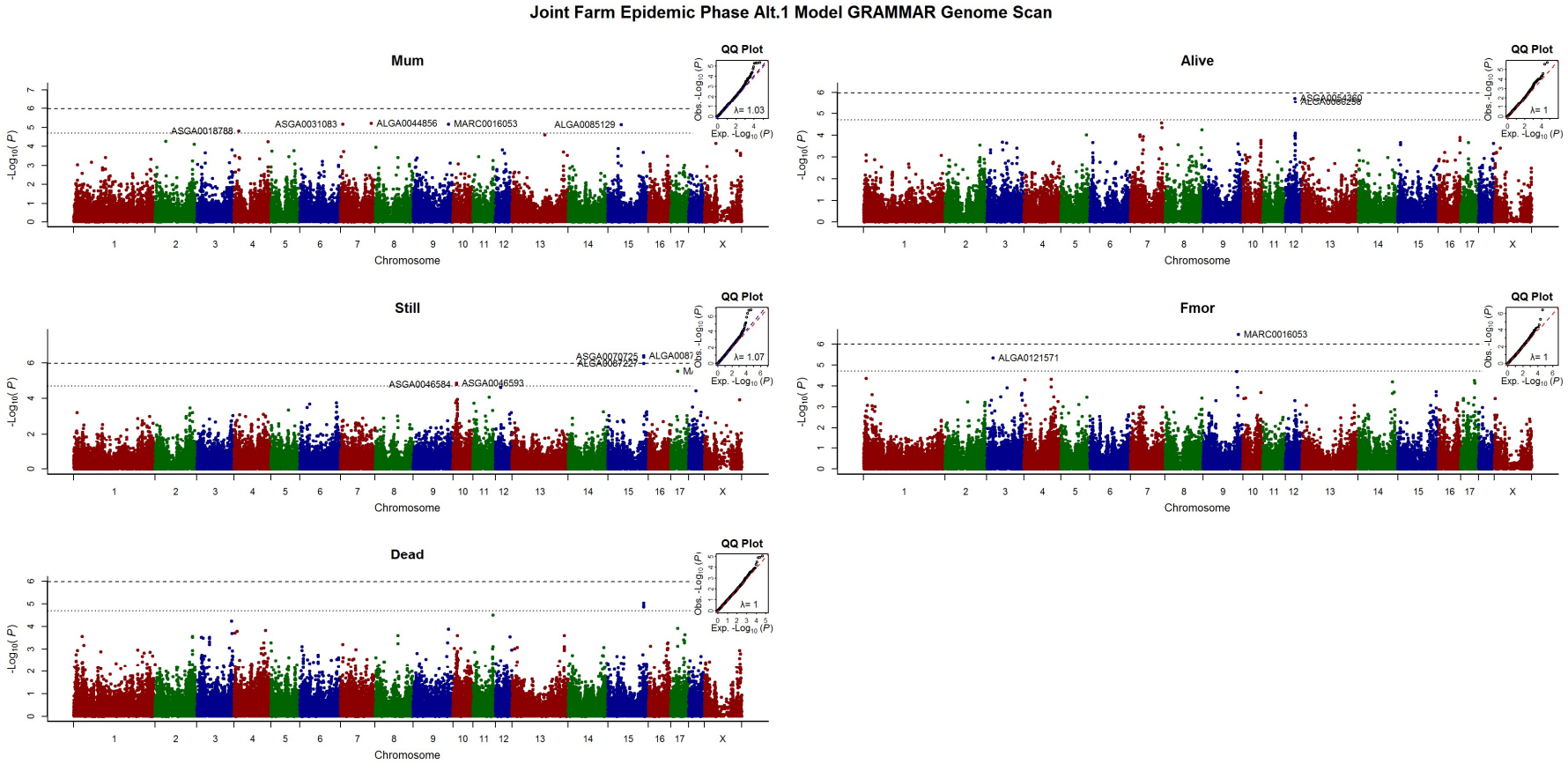

Figure 5.12 – Joint Farm Manhattan Plots using GRAMMAR for All Disease Indicator Traits Using the Alt.1 Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,976 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

Across all the reproductive traits considered, 14 SNPS show evidence of a significant effect above the chromosome level significance threshold, of which 3 SNPS show an effect at the genome wide threshold. 4 SNPs, including the 3 genome wide significant SNPS (MARC0016053, ALGA0087227, ALGA0087207 and ASGA0070725), appear in the results for more than one trait, and all of these are also indicated in the basic model analysis. Of the 14 SNPS, only 3 are unique to the Alt.1 model results, whereas 11 SNPs are common to the basic and Alt. 1 model genome scans.

Alive, *Fmor* and *Dead* show no evidence of inflation (λ=1) while *Still* and *Mum* show limited inflation in the test statistic at λ=1.03 and λ = 1.07, though no systematic change is seen to basic model estimates. GC is applied to the reported P values where λ>1. In all genome scans, a positive deviation in a seen in the observed $-\log_{10}(P)$ values for a few results at the upper end, as compared to that expected under a null distribution.

Using alternative model Alt. 2 a significant additive genetic effect could only be demonstrated for the *Still* trait.

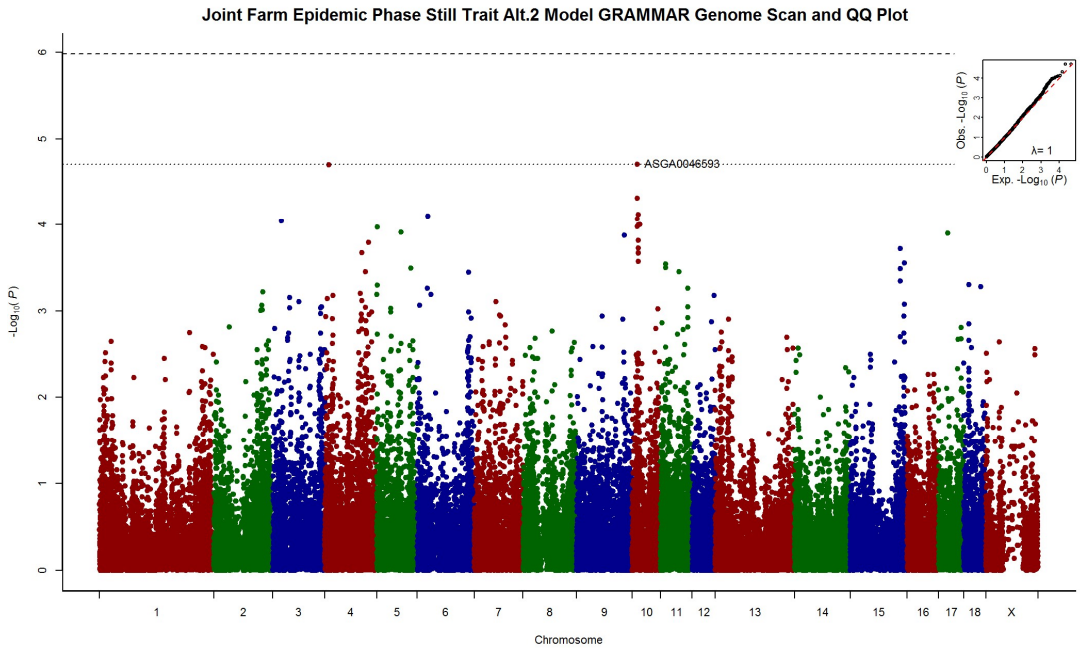

Figure 5.13 – Joint Farm Manhattan Plots using GRAMMAR for Still Trait Using the Alt.2 Model with Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,976 SNPs. SNPs significant above the indicative level shown labelled. QQ plots shows x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

Only one SNP (ASGA0046593) is indicated at the chromosome level of significance in this analysis (Figure 14). The same SNP was indicated for the basic and alternative model Alt 1 (Figures 12, 13).

In summary, in the models and traits presented for epidemic phase data (which omits Alt.2 model for the traits *Mum, Still, Dead* and *Fmor*) 16 unique SNPS are identified, some indicating an effect in several models and traits. The chromosome, position, p-values, reported effect size using GRAMMAR (including. SE) and MAF for SNPs showing an effect above the chromosome level of significance is shown in the appendix in Table A.11 – P Values and Effect Size Reported Using GRAMMAR for SNPs Significant at the Chromosome Level for Joint Epidemic Phase Analysis. A reduced table showing only unique Trait/SNPs combinations is shown in Table 5.6, where a SNP indicates an effect in

more than one model, the record with the lowest SNP P-value is shown, with additional models the SNP shows an effect using indicated.

Table 5.6 – P Values and Effect Size Reported Using GRAMMAR for Unique SNPs Significant Above Chromosome Level for Joint Epidemic Phase Analysis

Trait	Model	SNPS	SSC	Position	P-value	Effect Size	MAF	Other Models
Fmor	Basic	ALGA0121571	3	24,409,351	3×10^{-6}	0.12 (0.03)	0.05	Alt.1
Mum	Alt.1	ASGA0018788	4	18,736,505	2×10^{-5}	-0.1 (0.02)	0.14	
Mum	Basic	ASGA0031083	7	9,690,196	5×10^{-6}	0.08 (0.02)	0.43	Alt.1
Mum	Alt.1	ALGA0044856	7	118,565,139	6×10^{-6}	-0.09 (0.02)	0.14	
Mum	Basic	MARC0016053	9	135,725,781	2×10^{-7}	0.12 (0.02)	0.14	Alt.1
Fmor	Alt.1	MARC0016053	9	135,725,781	4×10^{-7}	0.06 (0.01)	0.14	Basic
Still	Basic	ASGA0046584	10	14,519,759	1×10^{-5}	-0.07 (0.01)	0.36	Alt.1
Still	Alt.1	ASGA0046593	10	14,647,576	1×10^{-5}	-0.07 (0.01)	0.36	Basic, Alt.2
Still	Basic	ASGA0053674	12	19,418,067	9×10^{-6}	0.07 (0.02)	0.31	
Alive	Alt.1	ASGA0054360	12	36,158,023	2×10^{-6}	-0.95 (0.2)	0.23	Basic
Alive	Alt.1	ALGA0066256	12	38,004,200	3×10^{-6}	-0.88 (0.19)	0.25	Basic
Mum	Alt.1	ALGA0085129	15	50,703,404	7×10^{-6}	-0.09 (0.02)	0.22	
Still	Alt.1	ALGA0087227	15	138,208,768	1×10^{-6}	-0.09 (0.02)	0.22	Basic
Still	Alt.1	ALGA0087207	15	138,593,445	4×10^{-7}	-0.09 (0.02)	0.23	Basic
Dead	Alt.1	ALGA0087207	15	138,593,445	1×10^{-5}	-0.11 (0.03)	0.23	
Still	Alt.1	ASGA0070725	15	138,715,882	5×10^{-7}	-0.09 (0.02)	0.22	Basic
Dead	Alt.1	ASGA0070725	15	138,715,882	1×10^{-5}	-0.11 (0.03)	0.22	
Still	Alt.1	MARC0016887	17	25,600,780	3×10^{-6}	0.15 (0.03)	0.06	Basic
Still	Basic	ALGA0097620	18	27,442,071	2×10^{-5}	0.12 (0.03)	0.09	

P-value corrected for λ where $\lambda>1$, Effect size is as reported in the GRAMMAR score and therefore is expected to underestimate the true effect. Where the same SNP occurs in multiple models the model with the lowest P -value shown. SNPs significant above the genome-wide significance threshold are shown highlighted in green.

In these results 3 regions are indicated as showing an effect by more than one SNP in the same chromosome region. These regions are on SSC10 (2 SNPs *Still* Trait, Basic model), SSC12 (2 SNPS *Alive* Trait, Alt.1 model) and SSC15 (3 SNPS *Still* trait all models, two of which are significant at the genome wide level).

Linkage disequilibrium (LD) was therefore investigated in these regions, and the r^2- values as a heat density plot for all pairwise LD in the region are shown in Figure 5.14. Strong levels of LD are clearly visible between the SNPs identified in the GRAMMAR genome scan (labelled) as compared to other SNPs in the region.

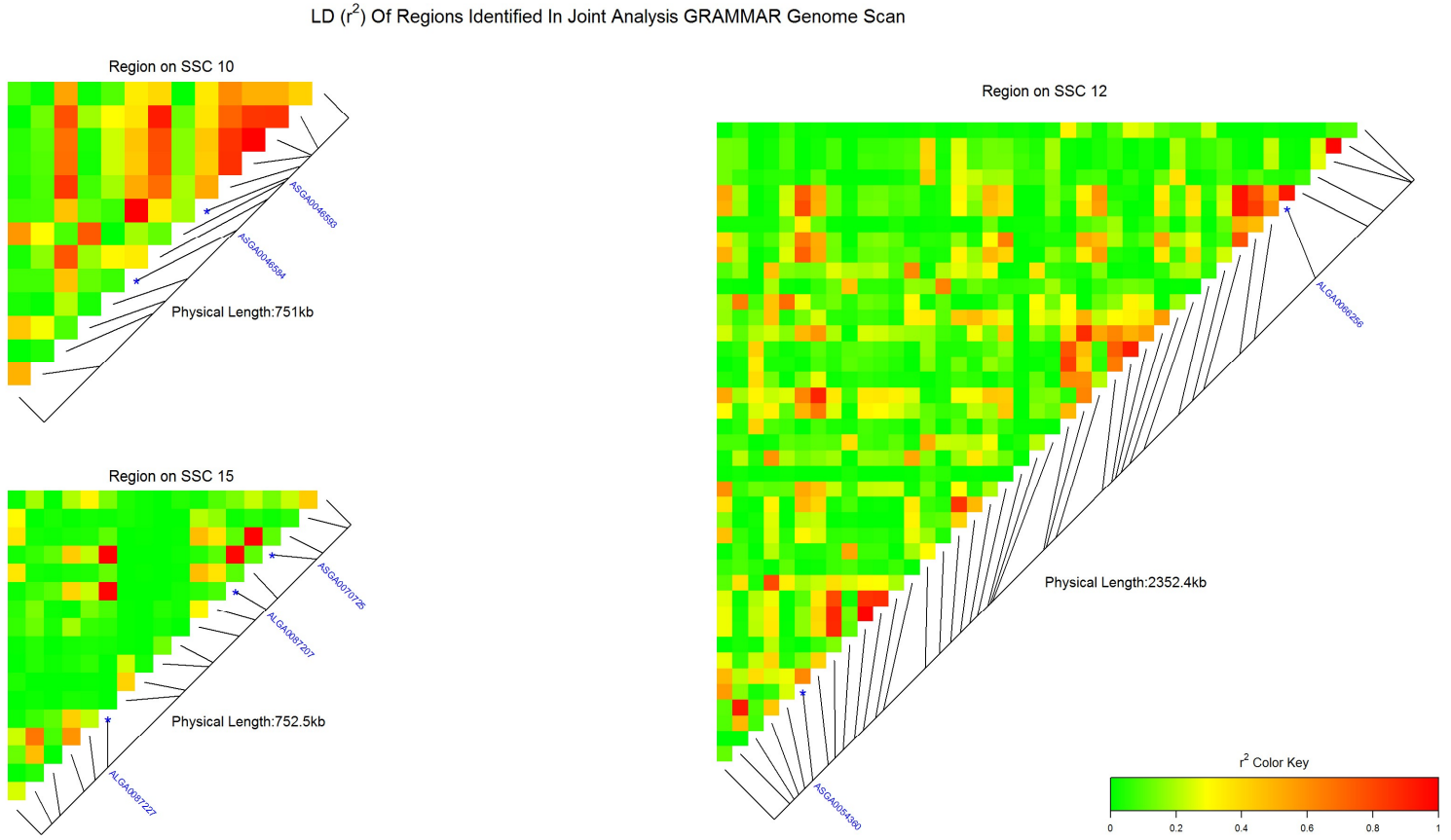

Figure 5.14 – LD (r2) of Regions Identified in Joint Analysis GRAMMAR Genome Scan

LD (r^2) shown for regions with more than one SNP significant above chromosome level (labelled) Region demarcated by the significant SNPs ± 5 SNPs provided for context.

The SSC15 region which contains 2 SNPs significant above the genome wide level show very strong LD (r^2=0.997 between the two SNPs indicated). Grouping the 7 SNPs into the 3 regions for which elevated levels of LD are observed between SNPs indicated in the genome scan, 13 unique regions (non-overlapping between traits and models and not indicating an effect in the same chromosomal region linked by LD) in nine different chromosomes indicate an association with reproductive performance traits in the epidemic phase joint farm analysis.

Four trait/model combinations were selected for the permutation analysis to estimate genome wide empirical significance based on the potential indicated in the presented genome scans. These were *Mum* and *Still* traits using the basic model and the *Alive* and *Fmor* traits using alternative model 1. With 10,000 times permutation genome-wide significance was estimated shown in Figure 5.15.

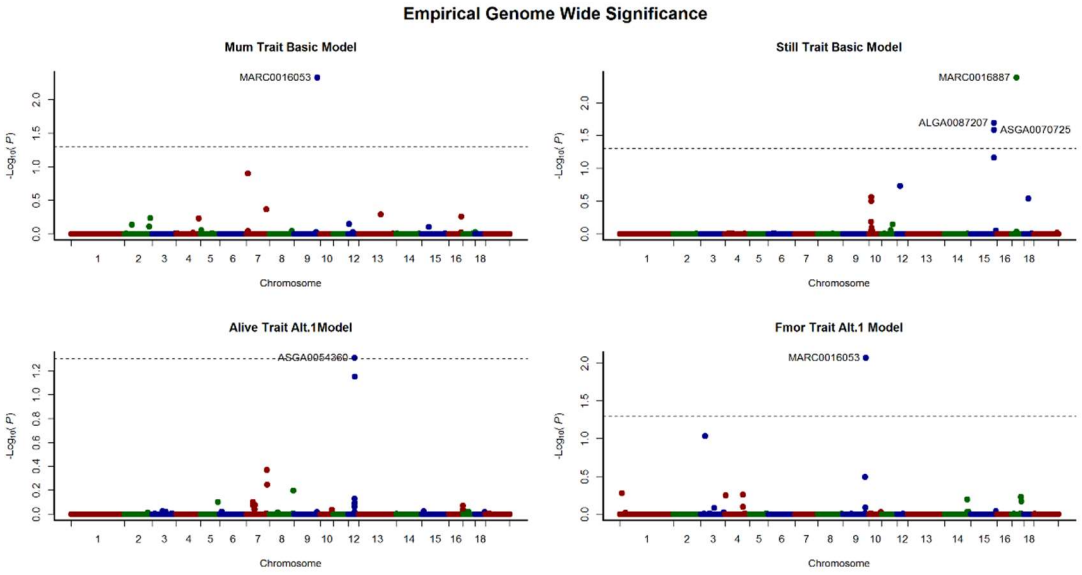

Figure 5.15 – Empirical Genome Wide Significance for Traits Suggesting Genome Wide Significant SNP Effects Using Basic and Alt.1 Models

Permutation analysis (×10,000) of trait and model combinations suggesting genome wide significant SNP effects using GRAMMAR analysis. Dotted line shows significance threshold at -$\log_{10}$(P=0.05).

Only five SNPs (MARC0016053, ALGA0087207, ASGA0070725, MARC0016887, ASGA0054360) show genome wide significance (P<0.05) in the permutation test. These five

SNPs were selected for further characterisation using the measured genotype approach to explore SNP effects, the relationship between variance components and the SNP effect and interactions between model terms and SNP effects, which could indicate confounding. One further SNPs was also included based on the evidence presented in the genome scans, namely ALGA0121571 on SSC 3 as it approaches genome wide significance for the *Fmor* trait using the basic model, appears above the chromosome significance threshold in the Alt.1 model and appears above the genome wide significance threshold in the Alt.2 model. Following permutation analysis on the *Fmor* trait model an empirical genome-wide p-value was estimated at P=0.092 (Alt.1 Model) and P=0.014 (Alt.2 Model).

Other Phases

GRAMMAR analyses were also carried out for the epidemic including unknown phase. Individual results are shown in the appendix for alternative model 1 (Figure A.9). In total 8 SNPs indicate significance in the joint epidemic including unknown phase analysis, of which 1 is unique and 7 also appear in the epidemic phase analysis (Table A.12 –). No new SNPS were found at the genome-wide significance threshold. Analyses using non-epidemic phase are shown in the appendix (Figure A.10). These were used to look for evidence of peaks in the same location as those presented, which were not seen.

Measured Genotype

In total, the following 6 SNPs that showed a significant (P<0.05) genome-wide effect in the above GRAMMAR GWA analyses, were considered for further characterisation using the measured genotype approach

- ALGA0121571 on SSC3
- MARC0016053 on SSC9
- ASGA0054360 on SSC12
- ALGA0087207 on SSC15

- ASGA0070725 on SSC15
- MARC0016887 on SSC17

Specifically, the individual SNPs were included as additional fixed effects in the linear mixed model analysis in ASREML for all reproductive traits in order to validate their significance and to estimate the allelic substitution effect and the percentage of additive genetic variance explained by the SNP in consideration. This was done using the basic model and alternative model 1, given the low additive genetic variance estimates and the potential therefore for confounding of the genetic effect for alternative model 2 was not considered for estimating effect sizes. Effect sizes are presented using the basic model with consideration made for Alternative model 1 to ensure that the effect on trait variance was significant when differences in epidemic were accounted for.

Using a smaller GRM, including only those animals included in the epidemic phase made no substantial difference to the variance component estimates. For comparison to the SNP models and calculation of the σ^2_{SNP} as a percentage of σ^2_A(Baseline) these estimates are provided in Table 5.7.

Table 5.7 – Variance Component and Heritability Estimates Generated for the Subset of Animals and SNPs in the Joint Farm Epidemic Phase Measured Genotype Analysis

Trait	σ^2_A (SE)	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum	0.03 (0.03)	0.56 (0.03)	0.06 (0.05)	0.05
Still	0.09 (0.03)	0.43 (0.02)	0.2 (0.06)	1×10^{-4}
Dead	0.05 (0.03)	0.65 (0.03)	0.07 (0.05)	0.04
Alive	0.87 (0.59)	10.9 (0.55)	0.08 (0.05)	0.03
Tof	0.92 (0.51)	8.8 (0.45)	0.1 (0.06)	0.01
Fmor	0.01 (0.01)	0.09 (5×10^{-3})	0.11 (0.05)	0.01

Variance components and heritability estimates generated using the basic, repeated measures model based on only joint farm epidemic phase animals shown in Table 5.2

To check whether the two SNPs on SSC15 showed any evidence of independent effects, both SNPs genotypes were fitted as fixed effects together in models for all traits. Only

ALGA0087207 was found significant (Wald F statistic P-values <0.05).As such SNP effects are only presented for ALGA0087207 on SSC15.

The results from this measured genotype analysis are shown in Table 5.8. As dominance effects were not found statistically significant for any of the traits / models, all SNP effects presented were calculated under the additive model.

Table 5.8 – P Values and Effect Size Reported Using Measured Genotype for Unique SNPs Significant At the Genome-Wide Level for Joint Epidemic Phase Analysis

SNP	Trait	Measured Genotype Results					SNP Effect Size Calculation			
		σ^2_A (SE)	σ^2_P (SE)	h^2 (SE)	LRT σ^2_A P	Wald P SNP	T-test P	α (SE)	σ^2_{SNP} (SE)	% σ^2_A
ALGA0121571	*Mum*	0.03 (0.03)	0.56 (0.03)	0.05 (0.05)	0.08	0.068	0.18	0.27 (0.22)	0.01	21.11%
ALGA0121571	*Still*	0.08 (0.03)	0.43 (0.02)	0.18 (0.06)	<0.001	0.059	0.10	0.31 (0.19)	0.01	11.05%
ALGA0121571	*Dead*	0.04 (0.03)	0.64 (0.03)	0.06 (0.05)	0.08	0.004	0.05	0.46 (0.23)	0.02	45.52%
ALGA0121571	*Alive*	0.71 (0.56)	10.69 (0.53)	0.07 (0.05)	0.05	<.001	0.07	-1.79 (0.98)	0.31	35.87%
ALGA0121571	*Tof*	0.95 (0.52)	8.82 (0.45)	0.11 (0.06)	0.01	0.677	0.27	-0.76 (0.86)	0.06	6.15%
ALGA0121571	*Fmor*	0.01 (5×10^{-3})	0.09 (5×10^{-3})	0.09 (0.05)	0.01	<.001	0.03	0.20 (0.09)	4×10^{-4}	38.11%
MARC0016053	*Mum*	0.03 (0.03)	0.55 (0.03)	0.05 (0.05)	0.11	<.001	5×10^{-4}	0.30 (0.08)	0.02	67.09%
MARC0016053	*Still*	0.09 (0.03)	0.43 (0.02)	0.2 (0.06)	<0.001	0.097	0.04	0.15 (0.07)	5×10^{-3}	6.79%
MARC0016053	*Dead*	0.04 (0.03)	0.64 (0.03)	0.06 (0.05)	0.09	<.001	9×10^{-4}	0.31 (0.09)	0.02	53.45%
MARC0016053	*Alive*	0.62 (0.57)	10.61 (0.53)	0.06 (0.05)	0.1	<.001	2×10^{-4}	-1.48 (0.38)	0.55	62.80%
MARC0016053	*Tof*	0.94 (0.52)	8.82 (0.45)	0.11 (0.06)	0.01	0.749	0.35	0.17 (0.34)	0.01	0.79%
MARC0016053	*Fmor*	0.01 (5×10^{-3})	0.09 (4×10^{-3})	0.09 (0.05)	0.02	<.001	5×10^{-4}	0.13 (0.03)	4×10^{-3}	40.41%
ASGA0054360	*Mum*	0.04 (0.03)	0.56 (0.03)	0.07 (0.05)	0.03	0.281	0.12	0.09 (0.06)	3×10^{-3}	8.84%
ASGA0054360	*Still*	0.08 (0.03)	0.43 (0.02)	0.19 (0.06)	<0.001	0.091	0.06	0.10 (0.05)	3×10^{-3}	3.94%
ASGA0054360	*Dead*	0.05 (0.03)	0.64 (0.03)	0.07 (0.05)	0.04	0.046	0.02	0.16 (0.06)	0.01	18.69%
ASGA0054360	*Alive*	0.91 (0.6)	10.87 (0.55)	0.08 (0.05)	0.02	0.069	0.04	-0.58 (0.27)	0.12	13.56%
ASGA0054360	*Tof*	0.84 (0.5)	8.68 (0.44)	0.1 (0.06)	0.01	0.002	0.02	-0.60 (0.23)	0.12	13.47%
ASGA0054360	*Fmor*	0.01 (0.01)	0.09 (5×10^{-3})	0.11 (0.05)	<0.001	0.086	0.03	0.05 (0.02)	1×10^{-3}	10.30%
ALGA0087207	*Mum*	0.04 (0.03)	0.56 (0.03)	0.06 (0.05)	0.04	0.05	0.03	-0.13 (0.06)	0.01	16.94%
ALGA0087207	*Still*	0.09 (0.03)	0.43 (0.02)	0.21 (0.06)	<0.001	0.355	0.16	-0.07 (0.05)	2×10^{-3}	1.79%
ALGA0087207	*Dead*	0.05 (0.03)	0.65 (0.03)	0.08 (0.05)	0.02	0.044	0.03	-0.14 (0.06)	0.01	15.11%
ALGA0087207	*Alive*	1 (0.6)	10.88 (0.55)	0.09 (0.05)	0.01	0.035	0.03	0.59 (0.26)	0.12	14.30%
ALGA0087207	*Tof*	0.99 (0.52)	8.82 (0.45)	0.11 (0.06)	0.01	0.427	0.20	-0.26 (0.23)	0.02	2.69%
ALGA0087207	*Fmor*	0.01 (0.01)	0.09 (5×10^{-3})	0.11 (0.05)	<0.001	0.085	0.04	-0.05 (0.03)	9×10^{-4}	8.97%
MARC0016887	*Mum*	0.03 (0.03)	0.56 (0.03)	0.05 (0.05)	0.07	0.005	0.34	-0.10 (0.18)	1×10^{-3}	3.01%
MARC0016887	*Still*	0.08 (0.03)	0.43 (0.02)	0.2 (0.06)	<0.001	0.048	0.29	0.12 (0.15)	2×10^{-3}	1.79%
MARC0016887	*Dead*	0.04 (0.03)	0.64 (0.03)	0.06 (0.05)	0.08	0.001	0.38	0.07 (0.20)	5×10^{-4}	1.05%
MARC0016887	*Alive*	0.64 (0.57)	10.66 (0.53)	0.06 (0.05)	0.08	<.001	0.13	-1.25 (0.83)	0.17	19.56%
MARC0016887	*Tof*	0.95 (0.52)	8.79 (0.45)	0.11 (0.06)	0.01	0.2	0.14	1.05 (0.73)	0.12	13.03%
MARC0016887	*Fmor*	0.01 (0.01)	0.09 (5×10^{-3})	0.1 (0.05)	0.01	0.005	0.23	0.08 (0.07)	7×10^{-4}	7.06%

Table 5.8 – P Values and Effect Size Reported Using Measured Genotype for Unique SNPs Significant At the Genome-Wide Level for Joint Epidemic Phase Analysis: α indicates estimate of the allele substitution effect using the additive model. Dominance models were not significant for any SNPs. σ^2_{SNP} provides estimate of SNP variance and $\%\sigma^2_A$ gives SNP variance as a percentage of the total baseline additive genetic variance (Table 5.7). Each trait and SNP previously identified as genome-wide significant shown shaded green

Using alternative model 1 the SNP Wald P values on trait variance were similar as those presented for the basic model. Models were also run to investigate the SNP effect in the presence of a SNP × population (farm and sow line) interaction. In all cases the SNP retained significance while the interaction term was not significant. This suggests that the model was adequately capturing the between farm and between sow line effect.

The only SNP which did not indicate a significant effect (P<0.1 Wald F test P) on the trait for which it was identified by the GWA analysis was ALGA0087207. However, a significant effect for this SNP was observed for all other disease indicator traits (p-values <0.085). The direction of the estimated allelic effect was consistent among the traits in consideration, i.e. a negative allelic effect for the *Alive* trait corresponded to a positive effect in loss traits (*Mum, Dead* and *Fmor*).

ASGA0054360 had a significant effect on the *Alive* trait, for which an association was detected in the GWA analysis, with a Wald F-statistic p value of 0.069. A significant effect for this SNP was also indicated in the *Still* (P=0.091), *Dead* (P=0.046) and *Fmor* (P=0.086) trait with significant (P<0.06) ASE. Again, the direction of estimated allele effects was consistent across the traits. This SNP is the only one which also shows an effect for the *Tof* trait (Wald P<0.002).

ALGA0121571 was not only confirmed to have a significant effect on *Fmor* trait variance using alternative Model 2 (as indicated by the GWA analysis) but was also significant when the basic or alternative model Alt.1 was used (P<0.01). A significant effect for this SNP was also found for the individual loss traits *Mum* and *Still (*P<0.068) and for both the *Dead* trait (P=0.004) and *Alive* trait (P<0.001).

The measured genotype approach not only confirmed the significant effect of MARC0016053 on *Mum* and *Fmor,* but also indicated a significant effect on the *Dead* and *Mum* traits (P<0.001).

No statistically significant interactions between SNP effects and other fixed effects (e.g. farm, sow line) were observed in any of the models. In summary, three SNPs were identified as showing a significant effect on the *Mum* trait (Wald P<0.05), i.e. MARC0016053, ALGA0087207 and MARC0016887, with MARC0016053 explaining the majority of the observed genetic variation. When all three SNPs were fitted simultaneously in the mixed model, only MARC0016053 and MARC0016887 were found to have a significant association with *Mum* (Table 5.9).Only one SNP, i.e. MARC0016887 was found to have a significant effect on the *Still* trait, explaining 1.79% of the variation.

All five examined SNPS were shown to be associated with the *Dead* trait, with each individual SNP explaining between 15-53% of the total additive genetic variance, except for MARC0016887, which only explained a negligible 1% of the total additive genetic variance. When all five SNPs were fitted simultaneously in the mixed model, only ALGA0121571, MARC0016053 and MARC0016887 remained significant (Table 5.9).

Similarly, when fitted individually, all of the examined SNPs, except for ALGA0054360, were found to have a significant effect on the *Alive* trait. Simultaneous fitting of these SNPs reduced the significant SNPs to the same three (ALGA0121571, MARC0016053 and MARC0016887) as found for the *Dead* trait (Table 5.9).

Finally, after both, individual and simultaneous fitting, the same three SNPs (ALGA0121571, MARC0016053, MARC0016887) demonstrated a significant effect on the *Fmor* trait.

Finally, to consider the effect these SNPs combined have on additive genetic variance, all SNPS indicating a significant effect on the variance of each trait were included as separate

fixed effects using the basic model. Variance component estimates and the SNP Wald statistic P value for these models are shown in Table 5.9

Table 5.9 – Variance Component Estimates and Wald P Value for Measured Genotype Model Including SNPs Previously Indicating an Assocation With Each Trait

Estimate		*Mum*	*Still*	*Dead*	*Alive*	*Tof*	*Fmor*
Variance Component Estimates	σ^2_A	0.03 (0.03)	0.08 (0.03)	0.03 (0.03)	0.34 (0.53)	0.84 (0.5)	0.01 (5×10^{-3})
	σ^2_P	0.55 (0.03)	0.43 (0.02)	0.62 (0.03)	10.19 (0.51)	8.68 (0.44)	0.09 (4×10^{-3})
	h^2 (SE)	0.05 (0.05)	0.2 (0.06)	0.04 (0.05)	0.03 (0.05)	0.10 (0.06)	0.07 (0.05)
	LRT P σ^2_A	0.11	2×10^{-4}	0.18	0.26	0.01	0.06
Wald F Statistic P value	ALGA0121571	–	–	0.003	0.001	–	0.001
	MARC0016053	0.001	–	0.002	0.001	–	0.001
	ASGA0054360	–	–	0.170	–	0.002	–
	ALGA0087207	0.083	–	0.135	0.063	–	–
	MARC0016887	0.010	0.048	0.005	0.001	–	0.011

– denotes not fitted.

For the *Mum, Dead* and *Alive* traits a significant heritable component cannot be demonstrated with the SNPs included. This suggests that these SNPs explain a considerable proportion of the additive genetic variance. Considerable reductions are seen in both additive genetic variance and heritability. A very marginal reduction is seen in additive genetic variance for the Still trait with MARC0016887 fitted (σ^2_A=0.08) as compared to the baseline (σ^2_A=0.09). This may be expected given the SNP variance as a percentage of total additive genetic variance was only 1.79%. While no reduction is seen in the additive genetic variance for the *Fmor* trait this is likely to be a result of rounding error given a reduction is seen in heritability from 0.11 in the baseline as compared to 0.09 with the three SNPs fitted.

To consider whether an effect could be demonstrated for these SNPs in non-epidemic phase the measured genotype analysis was performed using non-epidemic phase data. This considered the effect of each of the five SNPs on trait variance, fit separately for each SNP /

trait combination. No SNPs show evidence of a significant effect on trait variance in non-epidemic phase.

5.04 Discussion

Five SNPs have been identified and validated to have a significant effect on reproductive performance of sows during a PRRSV outbreak. Given that genetic variances for most reproductive traits during the non-epidemic phase were mostly negligible, and relatively small during the epidemic phase, estimated SNP effects were relatively large. There are a number of factors that could contribute to this. Power is a consideration in the experimental design of any GWA analysis. Whilst for the individual analyses and joint analysis sufficient power was indicated, these were not realised in terms of detectible genome wide significant SNPs. Whilst the trait could be polygenic and no detectible SNP effects are present, this is unlikely in light of the results of the joint analysis It is possible that incomplete exposure across the population reduced the power of the analysis (see Bishop & Woolliams 2010). In the absence of direct, individual measures of infection or disease status prevalence is not known and therefore the effect on power cannot be explored.

Also possible is the Beavis effect (Beavis, 1998). This describes an upward bias to the effect size calculated for SNPs included in the model as a result of low power in the model but which have an effect. The unexplained variance in the excluded SNPs may cause inflation of the effects of the SNPs included in the model. Finally, the lower heritability estimates observed in the SNP derived estimates of heritability as compared to the pedigree derived estimates of heritability (see chapter 3 discussion) may underestimate the total genetic variance, and thus lead to inflated SNP effects as a proportion of that variance.

Whilst some sensitivity is observed with respect to significance of SNPs when the alternative models are fitted (considering differences in epidemic and dynamic trend) some loci appear more sensitive than others. Only slight differences are seen between the models (Basic, Alt.1

and Alt.2), within farm and epidemic. The Manhattan profiles in the joint analysis more closely reflect the farm 1 analyses than the farm 2 analysis. However, our results clearly showed that combining both farms in the joint analyses is advantageous in term of statistical power to detect SNPs with genome wide significance.

With alternative model 1 fitted in the measured genotype analysis all the SNPs showed equally significant effects on trait variance as observed in the Basic model. Also tested was the effect of these SNPs on trait variance in non-epidemic phase where a significant effect ($P<0.1$) could not be demonstrated. Whilst this suggests that the observed SNP effects are specific to the PRRSV outbreaks, the reduced trait variance in non-epidemic phase could have limited the power to detect a significant effect.

Whilst it may appear self-evident that loci conferring benefits in terms the numbers born alive may confer reductions in terms of losses this may not necessarily follow. Indications are that the benefits conferred that the SNP effects analysed were consistent in terms of the direction of change in the loss (*Mum Still Alive Dead* and *Fmor*) and production (*Alive*) traits.

There are considerable differences between the results from the two single SNP GWAA methods employed, with considerably fewer SNPs identified using FASTA. However, these differences cannot be attributed to the methods themselves. The GRAMMAR method was employed using the environmental residuals from restricted maximum likelihood whereas FASTA was employed on the variance structures calculated using maximum likelihood, sometimes called full maximum likelihood (Kreft & de Leeuw, 1998). Using Restricted Maximum likelihood variance structures can be described as more realistic and larger than full maximum likelihood methods (Raudenbush & Bryk, 2002). Some similarities are noted, the genome-wide SNP indicated in the joint farm analysis on SSC15 is ~10Mbp from a SNP indicated for the Still trait FASTA Farm 1 analysis; the genome wide SSC9 SNP presented using GRAMMAR is 11Mbp from a SNP indicated using the FASTA analysis. Results from

preliminary scans using the GRAMMAR methodology on estimates using maximum likelihood estimates showed very similar results to the FASTA method with only minor changes to the reported P values but no change in terms of rank order.

Whilst genome scans are only shown for traits and models which show a significant heritable component, a couple of factors suggest that the corresponding environmental residuals could still be of value using the GRAMMAR analysis: firstly, all ASREML models converged and variance component estimates were not fixed at the lower boundary limit suggesting that, though non-significant, the additive genetic effect was removed in this process. Second, preliminary scans looking at the QQ plots from the GRAMMAR analysis on these residuals, inflation of the test statistic was not suggested (λ=1) and comparing the expected null distribution of the test statistic with the observed values no aberrations suggesting confounding are observed. Only a few instances of deviation from the line of unity are observed at the upper end of the distribution suggesting a few SNP effects detected in the results. These could be further considered at a later stage to explore other SNP associations.

Lewis et al. 2009 reported 11 SNPs significant at the genome-wide level in association with disease associated PRRS traits for farm 1 data using 4,595 SNPs. Given the small number of SNPs on the 7K SNP chip much lower genomic control parameters are set used (MAF< 0.01, Call Rate <40%) with 4,595 incorporated into the analysis. Of the reported significant associations 6/11 have a minor allele frequency of <0.05 which would be excluded in this analysis. One of the other differences between this analysis and the Lewis study is the inclusion of repeated measures and the fitting of a permanent environmental effect. Although the number of repeated records is limited, the actual estimates of permanent environmental variance are relatively high compared to the additive genetic variance estimates for the three traits (*Mum, Dead* and *Alive*) presented in the (Lewis *et al.,* 2009c) study. Given the size of this environmental effect, and the disproportionate contribution of records from the same animal, there is a risk of upward bias of the additive genetic variance component and SNP

effects. It is also possible that the higher number of SNPs in the Bonferroni method applied this analyses could have increased the conservativeness of the test in this analysis (Gao *et al.,* 2010).

Our results are in broad agreement with previous GWA studies of pigs. For example, the ALGA0087207 SNP identified on SSC15 lies between the 2 SNPs identified in (Yang *et al.,* 2016) in an analysis of PRRSV associated foetal death. ALGA0087207 is approximately 5 Mbp from each of the two previously demonstrate SNP associations demonstrated for MARC0055746 and ALGA0087932. No similarities are seen between these findings and those presented in Serão *et al.,* (2014). This may relate to the difference in the traits used. Whilst in PRRS phase one SNP is indicated on SSC1 for mortality traits the two QTL reported in SSC7 (further validated in Serão *et al.,* (2016)) are based on antibody titre, whether humoral response can be used as a resilience type trait is widely debated (Adamo, 2004).

Exploring the Animal Quantitative Trait Loci (QTL) Database (Hu *et al.,* 2005, 2016), several published regions in association with various traits span the position of the genome wide-significant SNPs identified in the joint farm analysis. A high number of phenotypes (between 39 and 61) are found at each locus. These contain a large number of carcass and meat quality traits which are of high economic value to the industry and are heavily funded topics of research. There are, however, a number of associations with disease or reproductive traits. For example, for ALGA0121571 (SSC3 at 24 Mbp) associations have been shown at this locus for both monocyte percentage (Gong *et al.,* 2010) at 22 Mbp – 38Mbp and the more general white blood cell count 22 Mbp – 67Mbp (Okamura *et al.,* 2012). Spanning the locus for MARC0016053 (SSC9 at 135 Mbp), associations shave been shown for *Salmonella* count in the spleen 126Mbp – 139Mbp (Galina-Pantoja *et al.,* 2009). For the locus identified on chromosome 12 (ASGA0054360 36Mbp) a number of disease susceptibility and immunological trait associations span this region *Actinobacillus pleuropneumoniae*

susceptibility 23Mbp-60Mbp (Reiner *et al.,* 2014); Interleukin 10 level, Toll Like Receptor (TLR) 9 level and TLR 2 level 24Mbp-60Mbp (Uddin *et al.,* 2011); and *Mycoplasma pneumonia* susceptibility 35-47Mbp (Okamura *et al.,* 2012). Spanning the SNP shown in association on SSC15 are several previously demonstrated associations with Salmonella count in liver and spleen 37-149Mbp (Galina-Pantoja *et al.,* 2009), Toll-like receptor 2 level (Uddin *et al.,* 2011), litter size and number born Alive (Schneider *et al.,* 2012). Finally, previously reported significant associations spanning the SNP locus identified on SSC18 (ALGA0097620, 27Mbp) include at 6Kbps-55Mbp *Actinobacillus pleuropneumoniae* susceptibility (Reiner *et al.,* 2014); 24-46Mbp *Mycoplasma hyopneumoniae* antibody titre (Uddin *et al.,* 2010); and 24-32Mps Interleukin 10 level (Uddin *et al.,* 2011). Some of these span large portions of a chromosome and the likelihood of finding an association covering a locus depends on this breadth of coverage. Whilst only the locus on SSC15 shows direct overlap with previously reported loci for PRRSV challenge traits, Yang *et al.,* (2016) also reported similar cross over between the QTL identified in his analysis and reported associations in the animal QTLdb for leukocyte profile, toll -like receptor expression and disease susceptibility traits.

Chapter 6. Regional Associations of Reproductive Traits Under PRRSV Challenge

6.01 Introduction

Regional Heritability Mapping (RHM) is an approach which has been suggested as a method of identifying genomic regions in association with traits, where individual SNP effects may not individually contribute enough variance to be detected (Nagamine *et al.,* 2012). This approach may help identify genomic regions associated with phenotypes measured in smaller, stratified populations, that could not be detected by single SNP GWAS (Riggio *et al.,* 2014). This methods has also been proposed as a means of accounting for the 'missing heritability' absent from some GWA analyses (Shirali *et al.,* 2016) and as a means of combing regional effects for improved statistical power (Uemoto *et al.,* 2013). In addition to analyses of human traits, it has also been used to explore the genomic architecture of nematode resistance in sheep (Riggio *et al.,* 2014). Given the low heritability estimates of reproductive performance traits obtained from SNP data in the previous chapter, and that only few SNPs with moderate effects on reproductive performance could be identified in the previous chapter, RHM seems a suitable method to investigate the genomic architecture underlying the reproductive performance data in this study.

This chapter will use Regional Heritability Mapping (RHM) to explore the distribution of the genetic variance in survival and mortality traits during PRRSV outbreaks across the genome. By dissecting the genome into windows and investigating the significance of the genetic trait variance explained by each window; the contribution of the window to the trait variance may be explored. To look for evidence of a putative SNP association in significant windows, each SNP in each significant window will be fitted individually as a fixed effect using the same model and data as used in the RHM, and the magnitude of the SNP effect will be calculated for statistically significant SNPs.

6.02 Materials and Methods

6.02.1 Data

The data used in these analyses is described in section 2.02.1. Specifically, only data from the two farms related to the ELISA confirmed PRRSV epidemic phase will be used in this chapter to identify regions of the genome which show an association with reproductive traits under PRRSV challenge. Furthermore, given an inability of the RHM methodology to fit a permanent environmental effect only data containing the first record per animal will be used, as described in section 5.02.1.

Of the 57,440 SNPs used in the GWA analyses, only the 52,648 mapped SNPs could be used in the RHM. SNPs and animals were subject to genomic quality control. 914 SNPs were excluded with a low call rate (<90%), 7,535 SNPs were excluded with a low MAF (<5%), 143 SNPs were excluded as they fell into both exclusion categories. Of the 1,553 animals for which genetic data was available 919 animals had records in the Epidemic Phase of which 31 were excluded because of low call rate (<90%), 3 were excluded because of too high autosomal heterozygosity (FDR<1%), 2 animals were excluded because of assumed errors in labelling and/or tracking of samples (corrected Identity-by-State (IBS) value > 1) leaving 883 animals for analysis.

For estimating the effect size of each putative SNP on reproductive performance phenotypes, the full genotype data was used. The QC applied to this data is shown in the 5.02.1 section: 'FASTA & Measured Genotype QC'.

6.02.2 Statistical Models

Response variables and fixed effects were the same as those used in the joint farm GWA analysis in Chapter 5. In particular, the three models (Basic, Alt.1 and Alt.2) were applied separately to each of the five disease indicator traits (*Mum, Still, Dead, Alive* and *Fmor*).

The regional heritability mapping methodology divides the genome into small consecutive overlapping regions. These regions, or windows, comprise a fixed number of SNPs for which a genomic relationship matrix (GRM) can be created (G_{WIN}). This is fitted in a full model including both the window genomic relationship matrix and the whole genome genomic relationship matrix (G_A), generated using all SNPs. This model is shown described in matrix notation in Equation 6.1.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_A u} + \boldsymbol{Z_W v} + \boldsymbol{\varepsilon}$$

Equation 6.1 – Matrix Equation of the Mixed Model

Where $\boldsymbol{y}$ is a vector of observations (the reproductive performance trait in consideration); $\boldsymbol{X}$ is a design matrix relating the observation to the fixed effects; $\boldsymbol{\beta}$ is a vector of fixed effects, $\mathbf{Z_A}$ is an incidence matrix of additive genetic random effects, $\mathbf{Z_W}$ is an incidence matrix of additive window random effects, $\boldsymbol{u}$ is a vector of additive genetic random effects, $\mathbf{v}$ is a vector of window additive genetic effects and $\boldsymbol{\varepsilon}$ is the error term. It was assumed that the random effects and errors were independently distributed: $u \sim N(0, \boldsymbol{G_A} \cdot \sigma^2_A)$; $v \sim N(0, \boldsymbol{G_{WIN}} \cdot \sigma^2_{WIN})$ and $\varepsilon \sim N(0, \boldsymbol{I} \cdot \sigma^2_E)$. $\boldsymbol{I}$ is an $n \times n$ identify matrix. The variance associated with the genomic relationships of the window (σ^2_{WIN}) are estimated, in addition to the estimate of genomic additive genetic variance (σ^2_A) and environmental (residual) variance (σ^2_E). From these values the total phenotypic variance ($\sigma^2_A + \sigma^2_{WIN} + \sigma^2_E = \sigma^2_P$), a genomic heritability ($\sigma^2_A / \sigma^2_P = h^2$) and a window heritability ($\sigma^2_{WIN} / \sigma^2_P = h^2_{WIN}$) can be calculated (Riggio *et al.,* 2014).

Furthermore, three types of GRMs were fitted, one where the whole population GRM is maintained as is, one where the between farm relationships are fixed at Zero and one where the between sow line relationships are fixed at zero. This method is applied in Riggio *et al.,* (2014) and also considered in Hess *et al.,* (2016) based on the work of Hayes *et al.,* (2009).

This considers relationships between populations as not fully known which may cause bias in the results. In total 9 RHM genome scans were run on each trait.

By comparing this full model to a reduced model, where the random window effect is omitted the significance of the random window effect was ascertained using a likelihood ratio test (LRT). The LRT statistic for the random window effect is assumed to be taken from a mixture of the X^2_0 and X^2_1 distributions (Self & Liang, 1987; Riggio *et al.,* 2014). A Bonferroni correction for the half the number of windows was applied to correct for number of independent tests. Significance was estimated at a genome wide level ($p<0.05$) and at an indicative level and that of one false positive per genome scan. The size of each window was 50 SNPs with a 25 SNP overlap between windows. Partitioning the data into 50 SNP windows at intervals of 25 SNPs resulted in 1,772 windows spanning the entire genome. From this a genome-wide significance threshold was calculated at 14.91 at the genome wide level or 10.60 at the indicative level for the LRT statistic, corresponding to P-values of 5.64×10^{-5} and 5.64×10^{-4} respectively. As in the previous chapter, QQ Plots were generated to investigate potential inflation and confounding.

6.02.3 Identification of Putative SNP Associations

Windows which were significant when the between sow line relationships are removed from the GRM were considered for further characterisation using the measured genotype approach. These were selected as they were considered more robust to population structure.

To investigate whether putative SNP associations could be identified in regions which indicate an effect, each SNP in each window that was significant at the suggestive level from the RHM analysis was fitted separately as a fixed effect in ASreml mixed model analyses. To ascertain the effect of the SNP in consideration on the window variance, the full model used in the RHM (including the window effect) was used alongside each SNP. To consider the effect of the SNP on the total additive genetic variance the reduced model from the

RHM, which omitted the window of the SNP, was also fitted alongside each SNP. To keep this directly comparable to the RHM results the same data was used. The between sow line genomic relationships fixed at zero in the window and whole genome GRM.

The significance of the SNP effects in the measured genotype models was ascertained using the Wald F Statistic from the reduced (omitting window effect) model. A Bonferroni corrected threshold of 0.05/50 = 0.001 was used to indicate a window-wide significant effect. The SNP was then checked to ensure it was significant ($P<0.05$) in the full model including the window effect. SNPs which show a significant effect on trait variance at the window wide level and show an impact on the heritable component described by the genomic relationships in the window (h^2_{WIN}) were considered putative associations.

6.02.4 Validation and Effect Size Calculation

Finally, these putative associations were characterised further with all traits using the full measured genotype method described. This was done using the full available phenotype data including repeated records by sow, fitting a permanent environmental effect and maintaining all relationships in the GRM. This calculates a SNP effect under the additive and dominance model (see Equation 5.3) and significance of the additive and Dominance allele substation effects assessed using a T-test. To check for interactions with population structure, separate models were also run to investigate the SNP effect in the presence of a SNP × population (farm and sow line) interaction.

Linkage disequilibrium (LD) was also used to explore linkage between any SNPs indicated on the same chromosome or within other regions which may be of interest.

All analyses were performed using ASREML 3.0 (Gilmour *et al.,* 2009). Plotting and data manipulation was done in R (R Core Team, 2016), GRM calculations were performed using GenABEL and LD investigated using the *r2fast* function (Aulchenko *et al.,* 2007b); LD plots were created using *LDheatmap* (Shin *et al.,* 2006).

6.03 Results

Given the slightly smaller SNP subset (including only mapped SNPs) a new set of epidemic baseline heritability estimates are shown Table 6.1. This table shows estimates for all three forms of GRM used: all relationships maintained, between farm relationships fixed at Zero and between sow line relationship fixed at Zero.

The results follow similar systematic patterns as described in Chapter 3 and Chapter 4. The estimates of heritability, with the cross-farm population fixed at Zero are consistently higher than with the whole population GRM maintained, except for the *Still* trait. The *Still* trait which has a moderate heritability estimate when the whole population GRM is maintained shows a slight reduction using the Basic model but increases in the Alt.1 and Alt.2 estimates. The estimates of heritability when the between sow line estimates are fixed at zero show consistently higher heritability estimates in all traits and models.

Table 6.1 – Epidemic Phase Baseline Heritability Estimates Using RHM Data

Trait	Model	GRM Maintained					Between Farm GRM Fixed at 0					Between Sow Line GRM Fixed at 0				
		σ^2_A (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	LRT P	σ^2_A (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	LRT P	σ^2_A (SE)	σ^2_E (SE)	σ^2_P (SE)	h^2 (SE)	LRT P
Mum	Basic	0.03 (0.03)	0.53 (0.03)	0.57 (0.03)	0.06 (0.05)	0.05	0.04 (0.03)	0.53 (0.03)	0.57 (0.03)	0.07 (0.05)	0.04	0.04 (0.03)	0.52 (0.04)	0.57 (0.03)	0.08 (0.06)	0.05
	Alt.1	0.03 (0.03)	0.53 (0.03)	0.56 (0.03)	0.05 (0.05)	0.08	0.03 (0.03)	0.52 (0.03)	0.56 (0.03)	0.06 (0.05)	0.06	0.03 (0.03)	0.52 (0.04)	0.56 (0.03)	0.06 (0.05)	0.1
	Alt.2	0.02 (0.02)	0.46 (0.03)	0.48 (0.02)	0.04 (0.05)	0.17	0.02 (0.02)	0.46 (0.03)	0.48 (0.02)	0.04 (0.05)	0.17	0.02 (0.03)	0.46 (0.03)	0.48 (0.02)	0.04 (0.05)	0.18
Still	Basic	0.09 (0.03)	0.34 (0.03)	0.43 (0.02)	0.2 (0.07)	3×10^{-4}	0.07 (0.03)	0.35 (0.03)	0.43 (0.02)	0.17 (0.07)	1×10^{-3}	0.09 (0.03)	0.34 (0.03)	0.43 (0.02)	0.21 (0.07)	3×10^{-4}
	Alt.1	0.07 (0.03)	0.35 (0.03)	0.42 (0.02)	0.18 (0.07)	1×10^{-3}	0.06 (0.03)	0.36 (0.03)	0.42 (0.02)	0.14 (0.07)	0.01	0.07 (0.03)	0.34 (0.03)	0.42 (0.02)	0.18 (0.07)	2×10^{-3}
	Alt.2	0.02 (0.02)	0.34 (0.02)	0.36 (0.02)	0.06 (0.06)	0.13	0.01 (0.02)	0.35 (0.02)	0.36 (0.02)	0.03 (0.05)	0.34	0.02 (0.02)	0.33 (0.02)	0.36 (0.02)	0.07 (0.06)	0.13
Dead	Basic	0.05 (0.03)	0.6 (0.04)	0.65 (0.03)	0.07 (0.05)	0.04	0.05 (0.04)	0.6 (0.04)	0.65 (0.03)	0.08 (0.05)	0.03	0.07 (0.04)	0.58 (0.04)	0.65 (0.03)	0.1 (0.06)	0.02
	Alt.1	0.03 (0.03)	0.59 (0.04)	0.63 (0.03)	0.06 (0.05)	0.08	0.04 (0.03)	0.59 (0.04)	0.63 (0.03)	0.06 (0.05)	0.08	0.04 (0.04)	0.58 (0.04)	0.63 (0.03)	0.07 (0.06)	0.07
	Alt.2	-	0.52 (0.03)	0.52 (0.03)	-	-	3×10^{-3} (0.02)	0.52 (0.03)	0.52 (0.03)	5×10^{-3} (0.04)	0.66	0.01 (0.02)	0.51 (0.03)	0.52 (0.03)	0.02 (0.05)	0.46
Alive	Basic	0.8 (0.6)	10.2 (0.68)	11 (0.56)	0.07 (0.05)	0.04	0.94 (0.62)	10.06 (0.69)	11.01 (0.56)	0.09 (0.06)	0.02	1.08 (0.71)	9.94 (0.74)	11.02 (0.56)	0.1 (0.06)	0.03
	Alt.1	0.58 (0.56)	10.14 (0.66)	10.72 (0.54)	0.05 (0.05)	0.1	0.62 (0.58)	10.1 (0.67)	10.72 (0.54)	0.06 (0.05)	0.08	0.7 (0.66)	10.03 (0.72)	10.73 (0.54)	0.07 (0.06)	0.1
	Alt.2	0.42 (0.45)	9.05 (0.57)	9.47 (0.47)	0.04 (0.05)	0.13	0.48 (0.47)	8.99 (0.58)	9.47 (0.47)	0.05 (0.05)	0.1	0.51 (0.53)	8.96 (0.61)	9.47 (0.47)	0.05 (0.06)	0.12
Fmor	Basic	0.01 (0.01)	0.08 (0.01)	0.09 (5×10^{-3})	0.1 (0.05)	0.01	0.01 (0.01)	0.08 (0.01)	0.09 (5×10^{-3})	0.11 (0.05)	4×10^{-3}	0.01 (0.01)	0.08 (0.01)	0.09 (5×10^{-3})	0.14 (0.06)	3×10^{-3}
	Alt.1	0.01 (5×10^{-3})	0.08 (0.01)	0.09 (5×10^{-3})	0.08 (0.05)	0.02	0.01 (5×10^{-3})	0.08 (0.01)	0.09 (5×10^{-3})	0.08 (0.05)	0.02	0.01 (0.01)	0.08 (0.01)	0.09 (5×10^{-3})	0.11 (0.06)	0.01
	Alt.2	2×10^{-3} (3×10^{-3})	0.07 (4×10^{-3})	0.07 (4×10^{-3})	0.03 (0.05)	0.24	3×10^{-3} (4×10^{-3})	0.07 (4×10^{-3})	0.07 (4×10^{-3})	0.04 (0.05)	0.18	4×10^{-3} (4×10^{-3})	0.07 (5×10^{-3})	0.07 (4×10^{-3})	0.05 (0.06)	0.15

LRT P values are shown for significance of the σ^2_A. The test statistic for the LRT calculated at twice the difference of the maximum likelihood with and without the σ^2_A included distributed under a mixture of $X_0^2$and X_1^2 distributions.

Heritability estimates were similar for the different GRMs, but slightly higher when between sow line relationships were fixed at Zero. Hence, to avoid repetition, only results for this GRM are presented here. RHM results using the other two forms of GRM fitted can be found in the appendix as indicated.

6.03.1 *Mum* Trait Analysis

Figure 6.1 shows the LRT for each window by its position on chromosome obtained from the *Mum* trait RHM genome scan, together with the corresponding QQ plots. SSC1 window 151 shows a peak across all models and is significant at the indicative level according to the models Alt.1 and Alt.2. The results are robust across the three types of GRMs (appendix Figure A.11).

The QQ plots confirm the presence of a significant genomic region for the *Mum* trait and show no evidence of confounding and no evidence of a more general pattern of inflation.

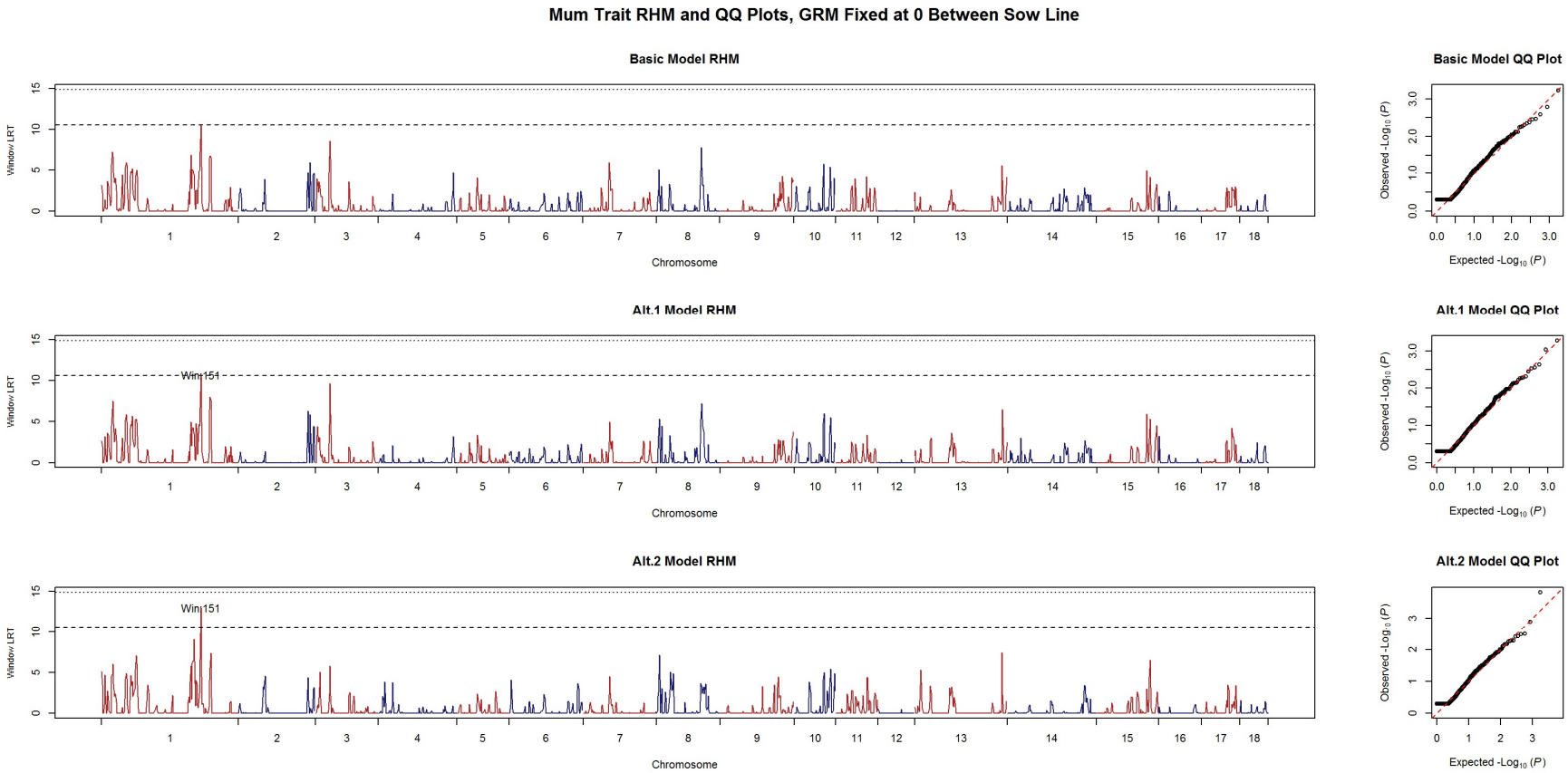

Figure 6.1 – Mum Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each of the 50 SNP windows at 25 SNP intervals, for *Mum* trait. Between sow line GRM was fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values plotted against those expected under a null distribution. The red dashed line refers to Y=X.

Similar to the findings for the single SNP genome scan methods some peaks appear sensitive to the effect of fitting the by epidemic effect and the trait trend in the alternative models. The peak indicated on SSC1 improves in significance between the basic, Alt.1 and Alt.2 models. However, this pattern is not consistent for all peaks.

The significant region; SSC 1, Window 151, is a 5.815Mbp region from ALGA0007541 (220,697,592) to ALGA0007602 (226,512,882). The window and genomic variance component estimates (σ^2_{WIN} and σ^2_A); heritability (h^2); window variance as a proportion of total phenotypic variance (h^2_{WIN}); and window variance as a percentage of baseline additive genetic variance, are shown in Table 6.2 for window significant above the indicative threshold in all models.

Table 6.2 – Mum Trait RHM Variance Component Estimates for SSC1 Window 151

Model	σ^2_{WIN}	σ^2_A	σ^2_E	h^2	h^2_{WIN}	Win LRT P	% σ^2_{WIN} / Baseline σ^2_A
Basic	0.05	0.04	0.51	0.07	0.09	6×10^{-4}¥	117%
Alt1	0.05	0.02	0.51	0.04	0.09	5×10^{-4}	167%
Alt2	0.05	0.01	0.45	0.02	0.1	2×10^{-4}	250%†

Win LRT P value significance thresholds are 5.64×10^{-5} at the genome wide and 5.64×10^{-4} at the indicative level. ¥denotes not significant at the indicative level, shown for reference.

Regional genomic variances were on generally higher than whole genomic variances. The σ^2_{WIN} remained constant throughout models, whereas estimates of σ^2_A differed between the models, with the smallest estimate for the Alt.2. model.

Following the measured genotype analysis of SSC1 window 151, no significant SNPs was found at the corrected threshold (Figure 6.2).

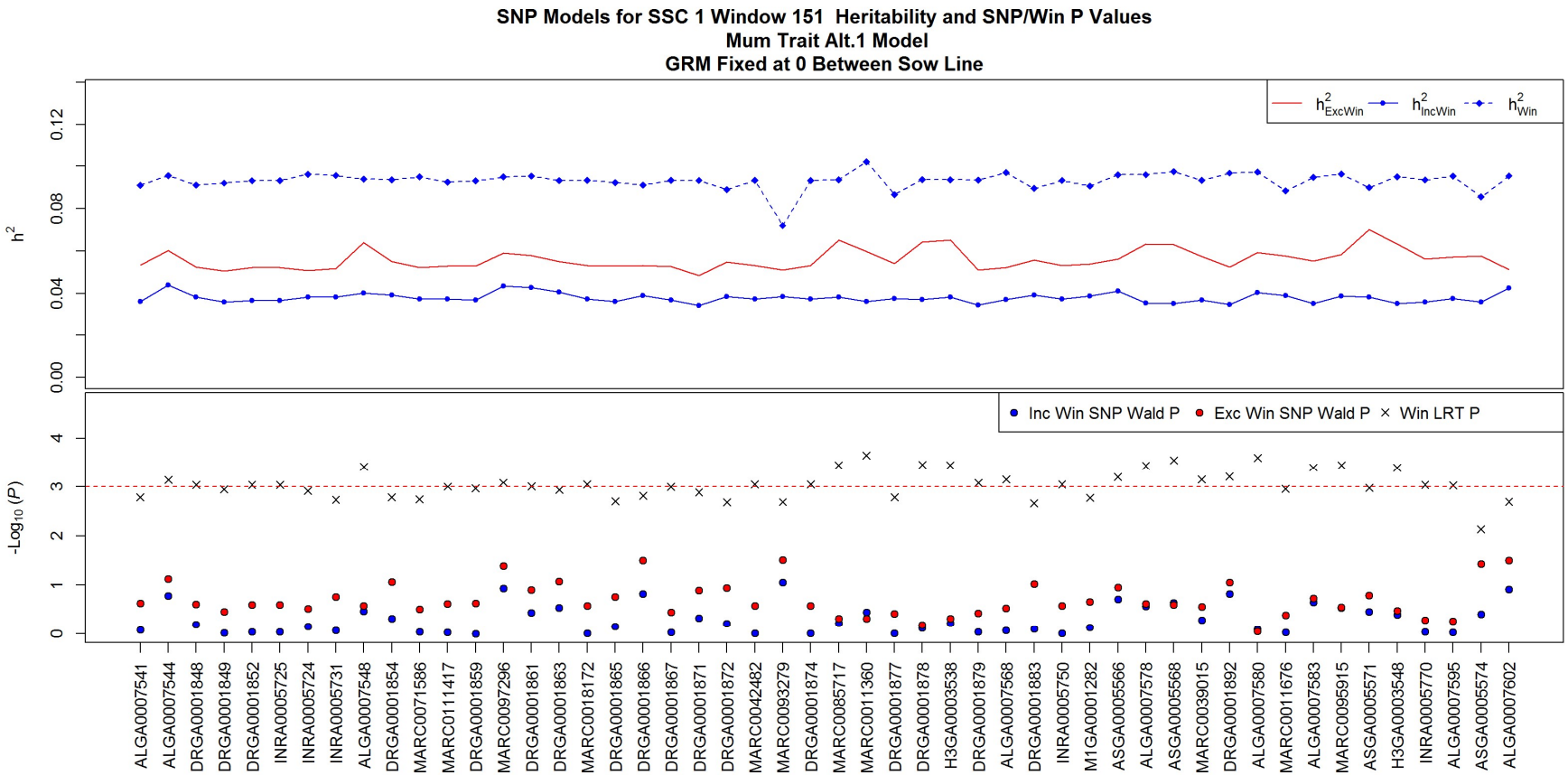

Figure 6.2 – Mum Trait, Alt.1 GRM Between Sow Line Fixed at Zero Analysis for SNPs in SSC1 Window 151 Showing Heritability and SNP/Window P-values.

SNP and Window estimates using the reduced model (red) the full model (blue). (Top) h^2 (solid lines) and h^2_{Win} (dashed line). (Bottom) -$\log_{10}$(P) values shown using the Wald F statistic for SNPs and LRT for the window. Bonferroni corrected P<0.05 value for 50 tests (red dashed line).

The top plot shows the estimates of whole genome heritability and window heritability with the SNP fitted in the full (including window effect) and reduced (omitting window effect) models. A slight reduction is seen in the window h^2 (blue dashed line) from the full model when SNP MARC0093279 is included (Figure 6.2 top). However, significance cannot be demonstrated at the window-wide level with the Wald P value (Figure 6.2 bottom), the actual Wald P values for this SNP are 0.088 and 0.031 in the full and reduced models respectively. Nor does fitting this SNP cause a considerable reduction in the significance associated with the window (×). An effect could be considered for ASGA0005574 with a marginal reduction in h^2_{WIN} a reduction in the significance of the window, and elevated levels of significance for the SNP, however, again, not at the corrected threshold. Given the nature of the RHM it is possible that this window indicates an effect due to the combined effects of SNPs. However, given the reduced threshold (one false positive per genome scan) at which this window effect is indicated at, it is also possible that no effect is present.

6.03.2 *Still* Trait Analysis

The *Still* trait plots of window LRT statistics for all three models fitted with genomic relationships fixed at Zero between sow line is shown in Figure 6.3. Across the models, five windows show significance at the indicative level in three chromosomes: SSC5, SSC6 and SSC9.

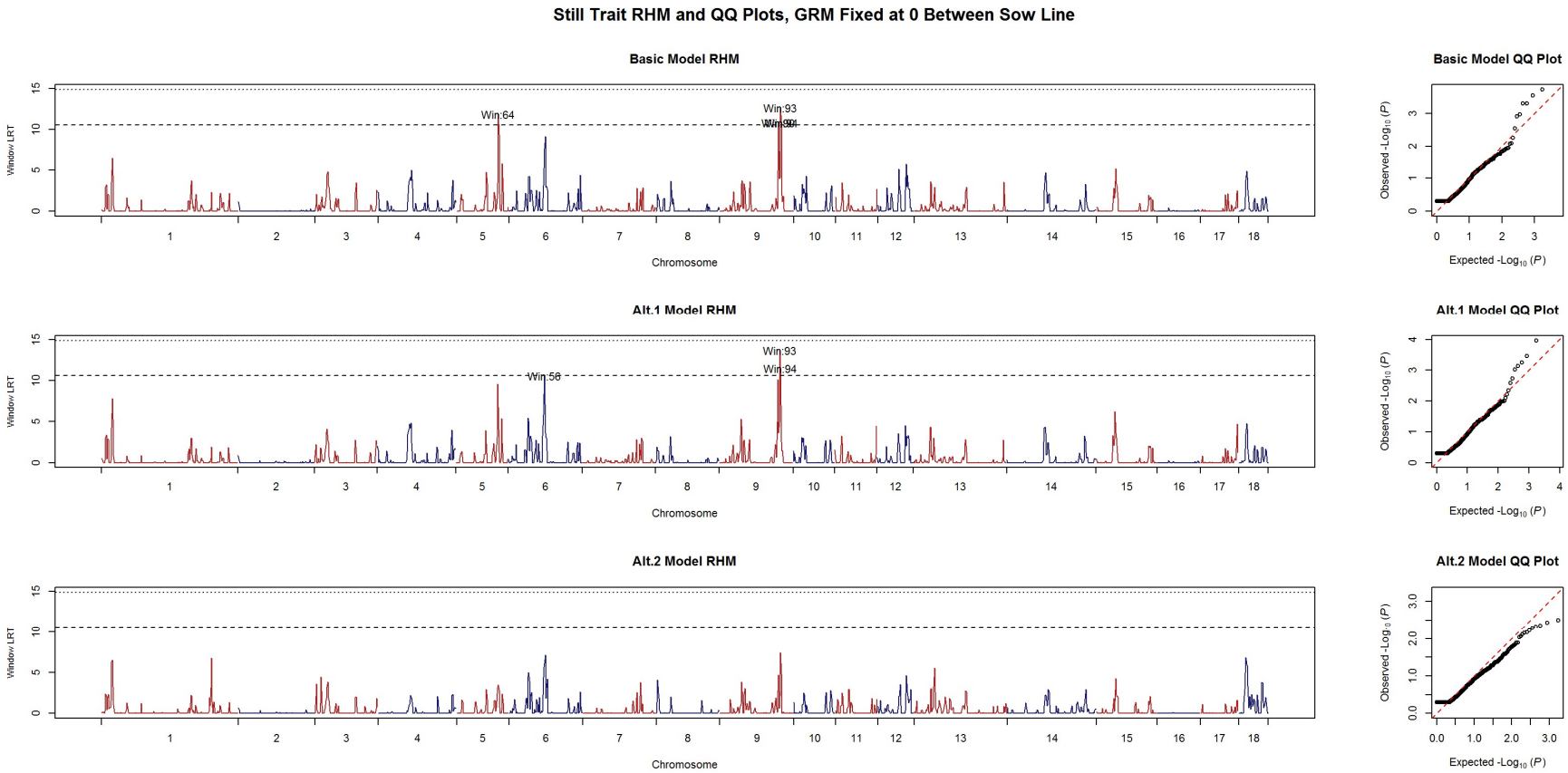

Figure 6.3 – Still Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

SNP and Window estimates using the reduced model (red) the full model (blue). (Top) h^2 (solid lines) and h^2_{Win} (dashed line). (Bottom) -$\log_{10}$(P) values shown using the Wald F statistic for SNPs and LRT for the window. Bonferroni corrected P<0.05 value for 50 tests (red dashed line).

The distribution of the observed test statistics compared to a null distribution; seen in the QQ plots, show no evidence for inflation in the test statistic or residual structure. There is strong evidence for a number of true window effects in the distribution of the test statistic for the Basic and Alt.1 models.

Window position and variance component estimates for windows significant above the indicative level in Figure 6.3 are shown in Table 6.3.

Table 6.3 – Still Trait RHM Variance Component Estimates for Windows Significant at the Indicative Level Across Models

SSC	Win	Start	End	Model	σ^2_{WIN}	σ^2_A	σ^2_E	h^2	h^2_{WIN}	Win LRT P	σ^2_{WIN} / $\sigma^2_{A\ Total}$
5	64	ALGA0033312 (89,832,719)	MARC0022126 (93,355,629)	Basic	0.03	0.07	0.33	0.16	0.07	3×10^{-4}	34%
6	56	ALGA0035761 (78,173,717)	ALGA0115725 (80,940,837)	Alt1	0.03	0.06	0.34	0.14	0.08	6×10^{-4}	47%
9	90	ALGA0054772 (127,310,801)	ASGA0102149 (130,798,057)	Basic	0.02	0.07	0.34	0.15	0.05	5×10^{-4}	26%
9	93	ALGA0123865 (132,656,946)	MARC0112726 (135,539,358)	Basic	0.03	0.06	0.34	0.14	0.08	2×10^{-4}	39%
				Alt1	0.03	0.05	0.34	0.11	0.08	1×10^{-4}	47%
9	94	ALGA0105037 (134,204,797)	MARC0072681 (136,813,512)	Basic	0.03	0.06	0.34	0.14	0.07	5×10^{-4}	33%
				Alt1	0.03	0.05	0.34	0.11	0.07	3×10^{-4}	40%

% σ^2_{WIN} /Baseline σ^2_A shown for the actual model estimates of σ^2_A regardless of significance (†P>0.1) Win LRT P value significance thresholds are 5.64×10^{-5} at the genome wide and 5.64×10^{-4} at the indicative level.

The window heritability estimates are consistently smaller than the whole genomic heritability and, in contrast to what was observed for the *Mum* trait, the resulting window variance as a percentage of baseline additive genetic variance are all <100%.

Two overlapping windows are seen significant at the indicative level on SSC9 N^{o}s 33 and 34, both these windows contain MARC0016053 demonstrated to have a genome-wide significant effect in the previous chapter.

In the measured genotype step, two SNPs in window 90 on SSC9 were significant at the corrected threshold (P<0.001) in the basic model see Figure 6.4.

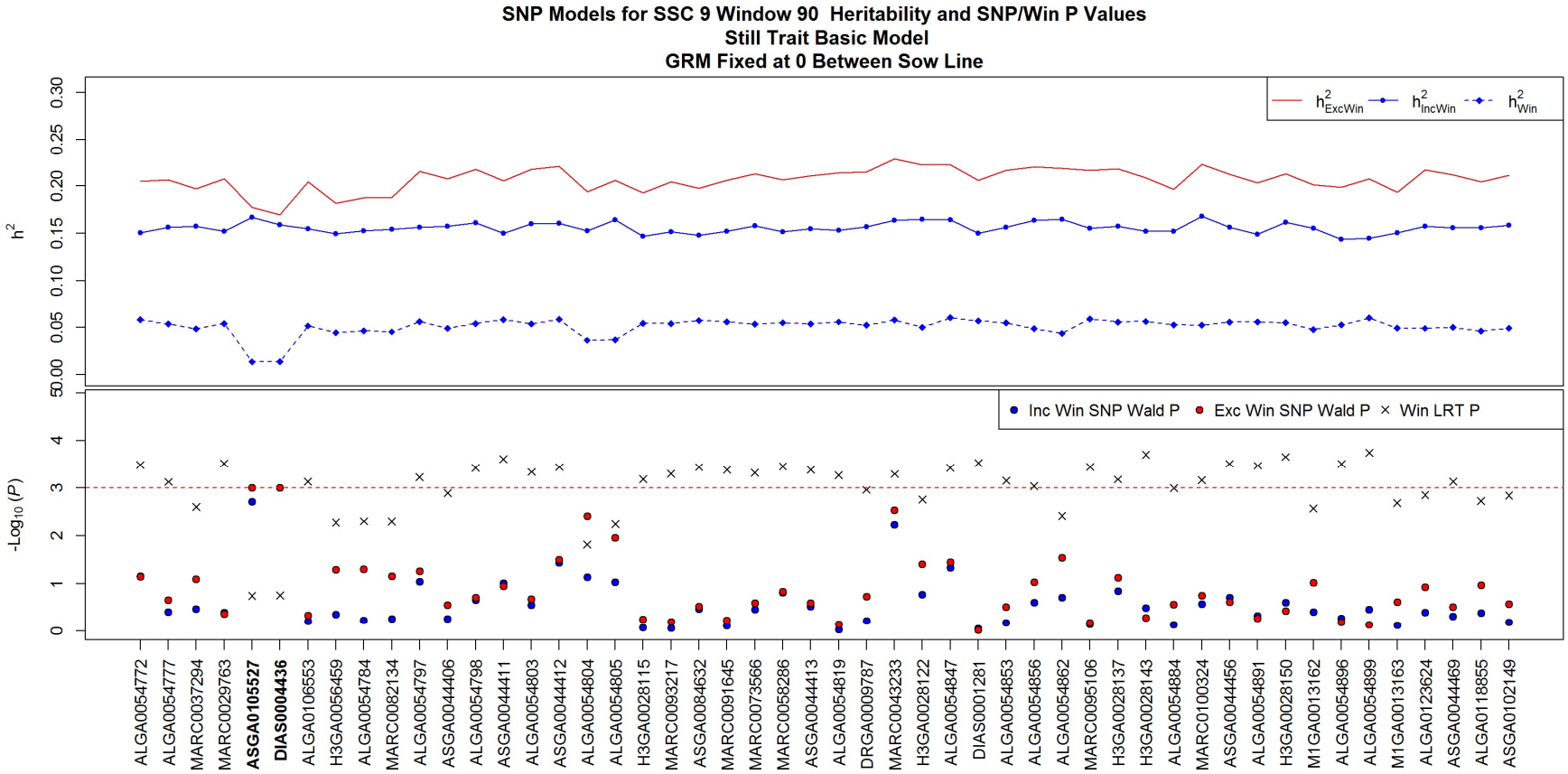

Figure 6.4 – Still Trait, Alt.1 GRM Between Sow Line Fixed at Zero Analysis for SNPs in SSC9 Window 90 Showing Heritability and SNP/Window P-values.

Estimates generated with the reduced model (red) the and full model (blue). (Top) Heritability estimates (solid line) and window h^2 (dashed line). (Bottom) $-\log_{10}(P)$ values shown using the Wald F statistic for SNPs and LRT for window. Minimum Wald P-value reported in ASreml is <0.001 (red dashed line), a SNP was considered significant when it met this threshold which is also the Bonferroni corrected P<0.05 value for 50 tests.

A reduction is seen in the window h^2 and whole genome heritability from the reduced model when either of two significant SNPs (ASGA0105527 and DIAS0004436, shown in bold in Figure 4) are included. Fitting either SNP reduces the window variance (σ^2_{win}) from 0.023 (s.e 0.011) to 0.006 (s.e 0.007). This would suggest that these SNPs account for a considerable proportion of the genomic variance explained by this region. Slight differences (below rounding thresholds) are seen between the estimates for the two SNPs suggesting they are not in perfect LD. These two SNPs were therefore taken forward for further characterisation and validation in section 6.03.6.

6.03.3 *Dead* Trait Analysis

Only one window (window 57 on SSC 10) indicates a significant effect on the Dead trait (Figure 6.5 and appendix Figure A.13), when model Alt2 was used.

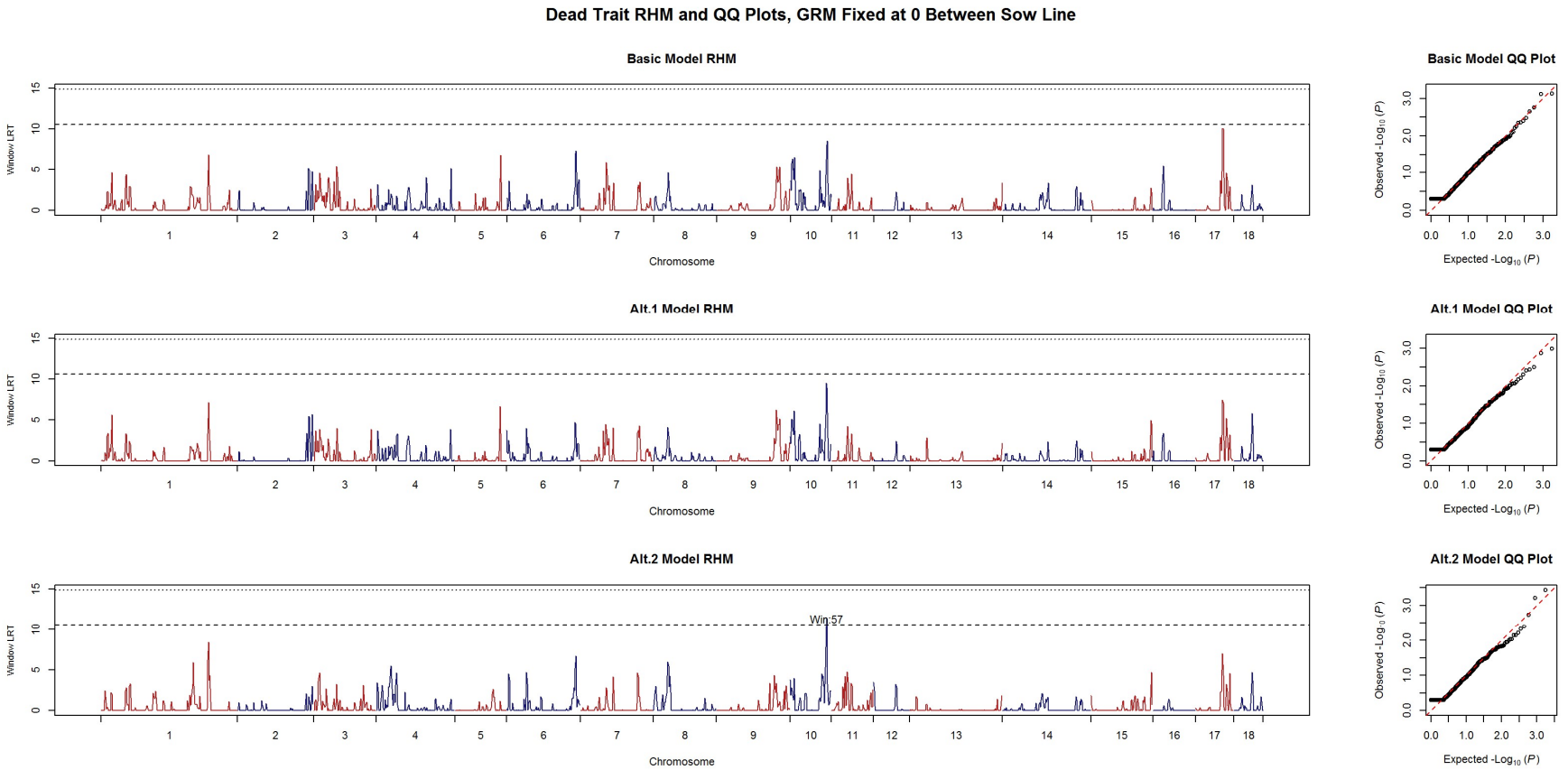

Figure 6.5 – Dead Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Dead* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

Slight confounding of the window effect at the upper end of the distribution is suggested in the QQ-plot by deviation below the line of unity for the Basic and Alt.1 models. In the Alt.2 model an effect is seen at the indicative level for SSC 10 Window 57 ($p=4\times10^{-4}$). An effect is also indicated in the corresponding QQ plot which shows no evidence of inflation or confounding. This window is also indicated in RHM genome scans where the whole population GRM is used, and with the Alt.2 model where the between farm GRM is fixed at Zero (appendix Figure A.13). In the basic model using the whole population GRM flanking windows 55 and 58 on this chromosome are also significant at the indicative level.

SSC 10 window 57 is a 1.466Mbp region from ASGA0105884 (69,064,005) to ASGA0048955 (70,529,789). The window variance was calculated at 0.03 and the whole genome additive genetic variance at 0.04, with corresponding heritability estimates of $h^2=1\times10^{-7}$ and $h^2_{Win}=0.05$.

In the measured genotype step (using the Alt.2 model), two SNPs (ASGA0083169 and MARC0020386) show a significant effect on the *Dead* trait variance at the corrected threshold (Wald P <0.001). Models fitting these SNPs also show reductions in the window heritability (h^2_{WIN}) (see Figure 6.4).

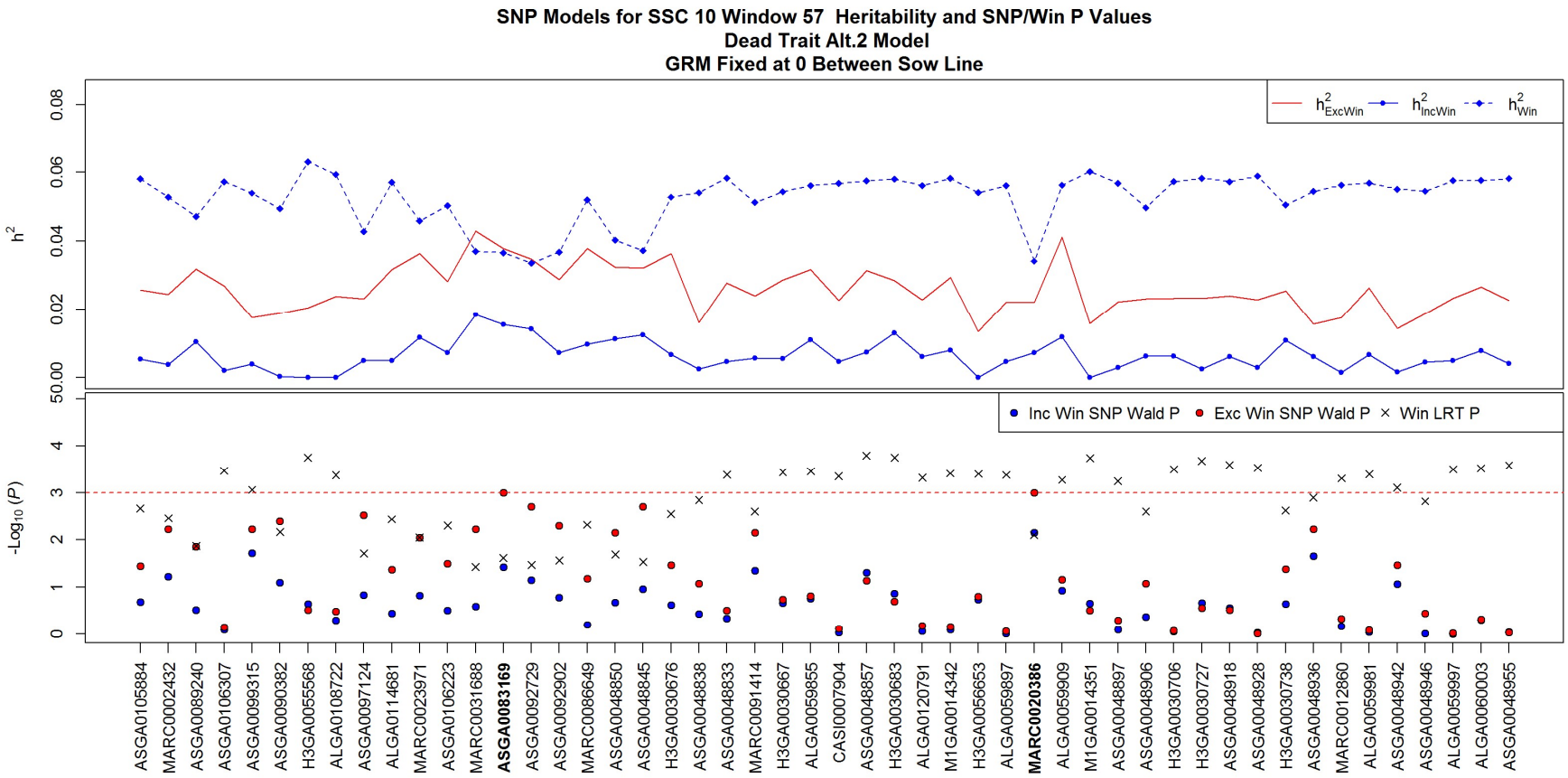

Figure 6.6 – Dead Trait, Alt.2 Model, GRM Between Sow Line Fixed at Zero SNPs Analysis for SSC10 Window 57, Heritability and SNP/Window P-values.

Estimates generated under the reduced model (red) the and the full model (blue). (Top) Heritability estimates (solid line) and, where fitted window h^2 (dashed line). (Bottom) $-\log_{10}(P)$ values shown using the Wald F statistic for SNPs and LRT for window.

In the full model (including the window effect) the actual σ^2_{WIN} estimates reduced from 0.029 (s.e 0.015) in the baseline to 0.019 (s.e. 0.013) as a result of fitting ASGA0083169; and to 0.018 (s.e 0.012) as a result of fitting MARC0020386 with corresponding SNP Wald P values of 0.001. This suggests that both SNPs capture a significant proportion of the regional genetic variance. These two SNPs were therefore taken forward for further characterisation and validation in section 6.03.6.

6.03.4 *Alive* Trait Analysis

Four non-overlapping regions indicate a significant effect at the reduced threshold across models, following RHM analysis using the *Alive* trait (Figure 6.7). Two regions (SSC3 window 23 and SSC4 overlapping windows 116 and 117) indicate an effect in the basic model, of which the region on SSC4 also indicates an effect in Alt.1 model (though only in window 117). SSC7 Window 89 is significant only using the Alt.1 model and SSC9 window 23 is only significant in the Alt.2 model.

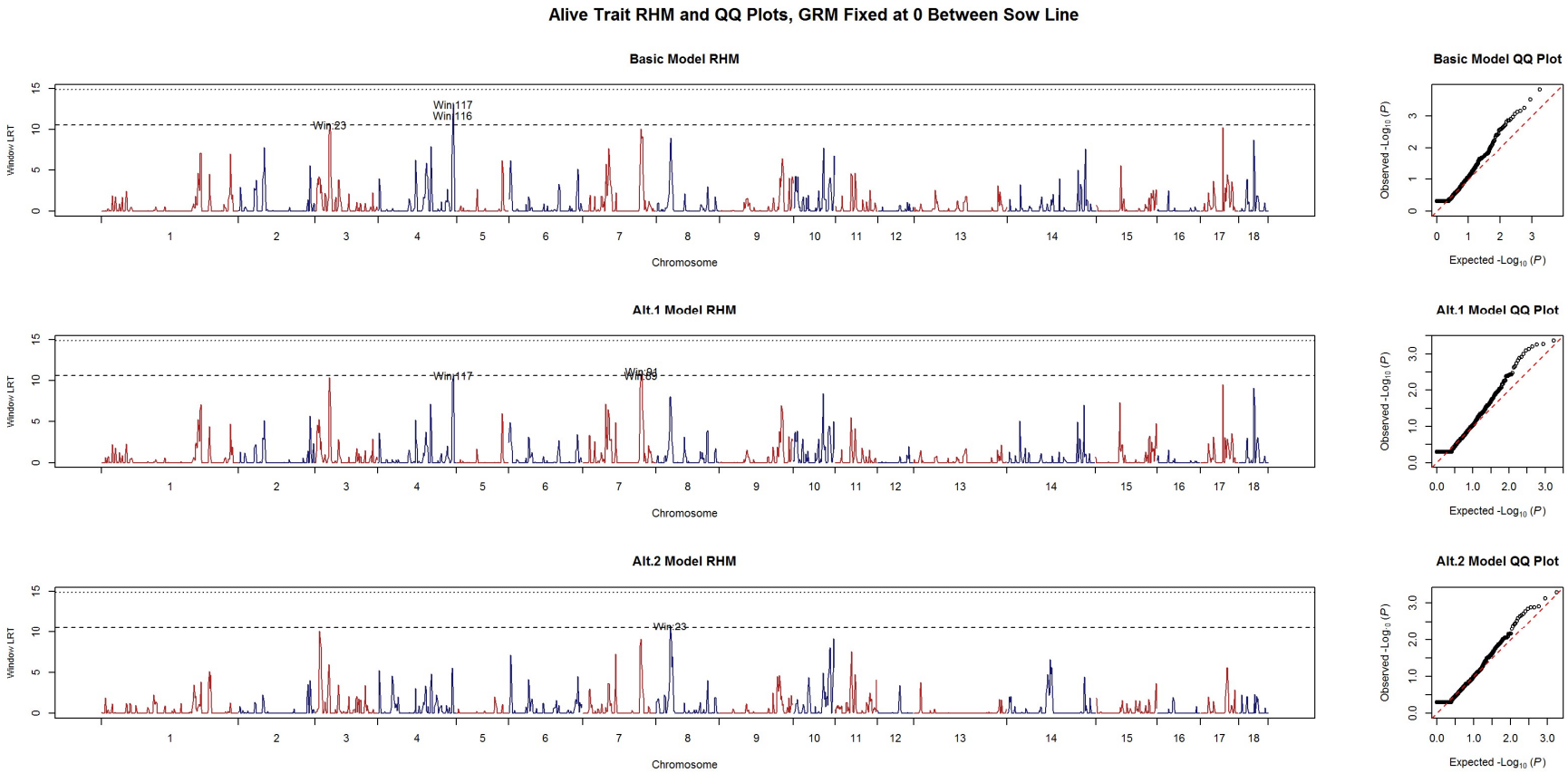

Figure 6.7 – Alive Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Alive* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

A slight irregular pattern is seen in the distribution of the test statistic in the QQ plots for Figure 6.7 which could indicate some confounding, inflation, or a high degree of genomic architecture relating to the trait. Similar patterns are seen for the results using the full GRM and that fixed at Zero between farm shown in appendix Figure A.15. The results where the between sow line population is fixed at Zero indicate greater control of any inflation than the other analyses.

Of the RHM genome scans reported in figure 9, the window variances for the six windows which indicate an effect on the *Alive* trait, across the models, are shown in Table 6.4.

Table 6.4 – Alive Trait RHM Variance Component Estimates for Windows Significant at the Indicative Level Across Models

SSC	Win	Start	End	Model	σ^2_{WIN}	σ^2_A	σ^2_E	h^2	h^2_{WIN}	Win LRT P	σ^2_{WIN} / $\sigma^2_{A\ Total}$
3	23	ASGA0104499 (28,886,788)	ALGA0018345 (30,846,173)	Basic	1.23	0.57	9.84	0.05	0.11	6×10^{-4}	114%
4	116	ALGA0029371 (137,290,094)	ALGA0029505 (139,182,345)	Basic	0.95	0.8	9.62	0.07	0.08	3×10^{-4}	11.47%
4	117	ASGA0023261 (138,291,968)	ALGA0029567 (140,216,348)	Basic	1.11	0.89	9.61	0.08	0.1	1×10^{-4}	103%
				Alt1	1.07	0.53	9.71	0.05	0.09	6×10^{-4}	152%
7	89	ASGA0035741 (107,618,108)	ASGA0035822 (110,686,764)	Alt1	0.57	0.29	9.92	0.03	0.05	5×10^{-4}	81%
7	91	DRGA0008061 (110,711,289)	INRA0028035 (113,153,185)	Alt1	0.61	0.35	9.86	0.03	0.06	4×10^{-4}	87%
8	23	ALGA0047027 (26,734,852)	DRGA0008477 (28,972,036)	Alt2	0.24	0.02	9.11	2e-03	0.03	5×10^{-4}	47%†

% σ^2_{WIN} /Baseline σ^2_A shown for the actual model estimates of σ^2_A regardless of significance (†P>0.1) Win LRT P value significance thresholds are 5.64×10^{-5} at the genome wide and 5.64×10^{-4} at the indicative level.

The measured genotype analysis showed that MARC0095146 in Window 91 of SSC7 was significant at the corrected threshold (P<0.001). The whole genome heritability estimate reduces from 0.07 (s.e 0.06) to 0.01 (s.e 0.06) when this SNP is included in the model. Similarly, h^2_{Win} drops from 0.06 (0.03) without the SNP to 0.03 (s.e 0.02) with the SNP included. These considerable reductions in heritability across all estimates with this SNP fitted are shown in appendix Figure A.15 which coincide with a reduction in the significance of the window. This SNP was therefore taken forward for further characterisation and validation in section 6.03.6.

6.03.5 *Fmor* Trait Analysis

The plot of window LRT following RHM using the *Alive* trait for all models are shown in Figure 6.8. Similar patterns are observed in both the Manhattan plots and QQ plot as for the *Alive* trait. The exact window locations, variance component estimates and LRT P values, for windows significant above the indicative threshold are shown in Table 6.5.

SSC4 Win 117 reaches genome wide significance using the *Fmor* trait, accounting for 80% of the total σ^2_A. This window encompasses the region reported in Boddicker et al. 2012 in association with viral load and weight gain. This region includes all 6 SNPs in perfect LD in the Boddicker study, found to capture 99.4% of the additive genetic variance in this region. SNP WUR1000125 showed the greatest association with trait variance across arrange of studies (Boddicker *et al.,* 2012, 2014b, 2014a; Hess *et al.,* 2016). The ~1.4Mbp region identified in Boddicker et al. (2012) is shaded, the 6 top SNPs are shown labelled on the LD plot and bold on the axis for single SNP models. The 6 SNPS reported in Boddicker et al. (2012) are not in perfect LD (r^2=0.89-0.99, μ= 0.94) in this analysis suggesting different patterns of linkage within this population. From the measured genotype analysis, no individual SNP shows a significant effect on *Fmor* trait variance nor reduced the trait h^2_{win} (Figure 6.9).

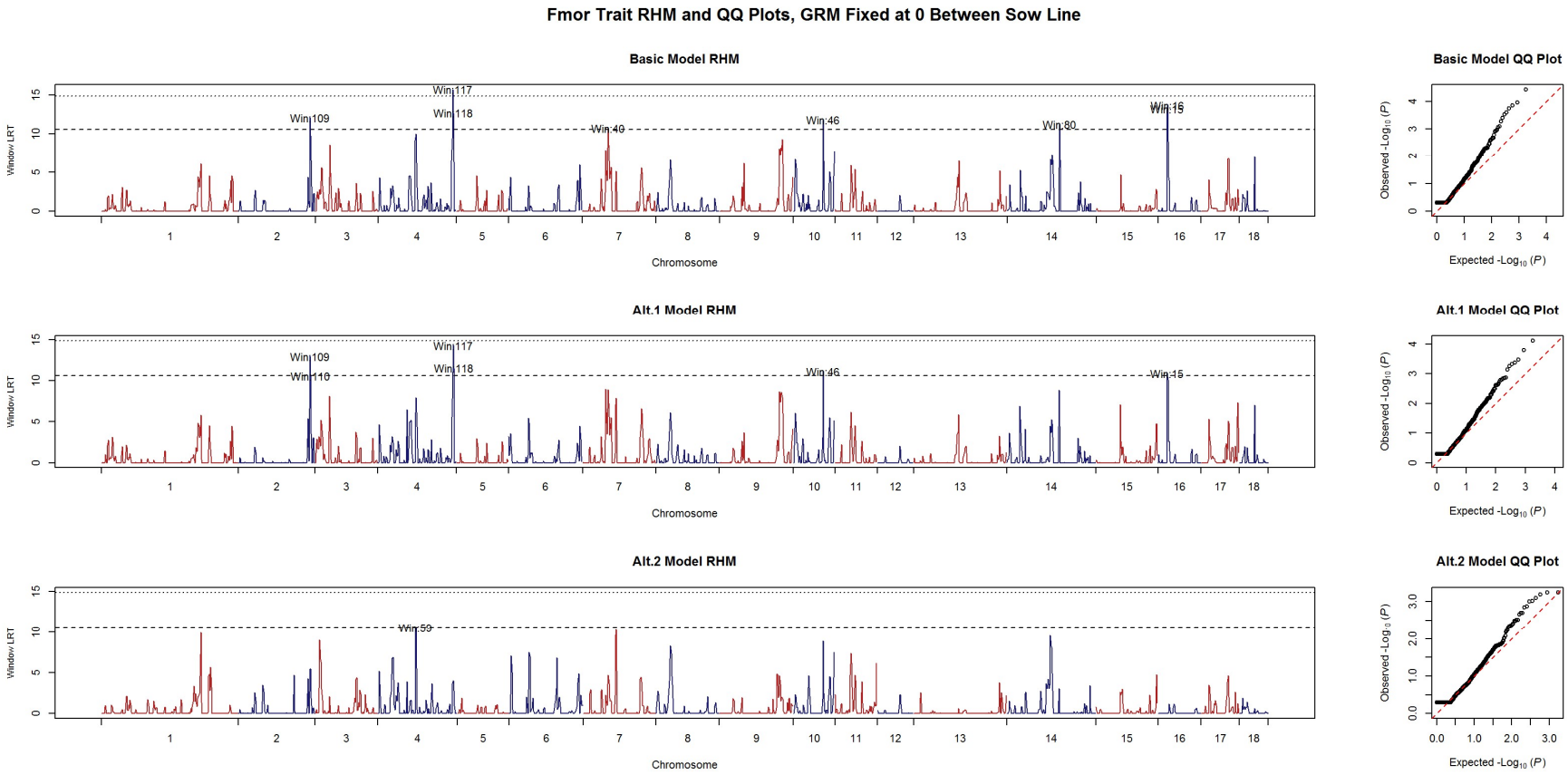

Figure 6.8 – Fmor Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Fmor* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

Table 6.5 – Fmor Trait RHM Variance Component Estimates for Windows Significant at the Indicative Level Across Models

SSC	Win	Start	End	Model	σ^2_{WIN}	σ^2_A	σ^2_E	h^2	h^2_{WIN}	Win LRT P	σ^2_{WIN} / $\sigma^2_{A\ Total}$
2	109	ALGA0016883 (152,941,065)	H3GA0054067 (154,466,689)	Basic	0.01	0.01	0.08	0.09	0.07	2×10^{-4}	51%
				Alt1	0.01	0.01	0.08	0.06	0.08	2×10^{-4}	75%
2	110	ASGA0095249 (153,920,242)	ALGA0122013 (155,198,988)	Alt1	3×10^{-3}	0.01	0.08	0.07	0.03	6×10^{-4}	30%
4	59	ASGA0020028 (74,573,226)	ALGA0025716 (76,762,171)	Alt2	0.01	3×10^{-3}	0.07	0.03	0.15	6×10^{-4}	327%
4	117	ASGA0023261 (138,291,968)	ALGA0029567 (140,216,348)	Basic	0.01	0.01	0.08	0.11	0.11	4×10^{-5}	80%
				Alt1	0.01	0.01	0.08	0.08	0.11	8×10^{-5}	109%
4	118	ASGA0023314 (139,260,673)	M1GA0006869 (141,552,372)	Basic	0.01	0.01	0.08	0.11	0.09	2×10^{-4}	67%
				Alt1	0.01	0.01	0.08	0.07	0.09	3×10^{-4}	92%
7	40	ASGA0033134 (44,021,990)	ASGA0033252 (46,959,202)	Basic	0.01	0.01	0.08	0.12	0.08	5×10^{-4}	60%
10	46	H3GA0055918 (57,417,764)	MARC0076652 (59,911,824)	Basic	0.02	0.01	0.08	0.06	0.18	3×10^{-4}	139%
				Alt1	0.02	3×10^{-3}	0.08	0.03	0.18	4×10^{-4}	186%
14	80	INRA0045306 (90,210,223)	ASGA0064890 (92,714,310)	Basic	0.01	0.01	0.08	0.13	0.08	4×10^{-4}	63%
16	15	ASGA0072476 (20,071,271)	ALGA0104683 (21,890,014)	Basic	0.02	0.01	0.08	0.1	0.22	1×10^{-4}	182%
				Alt1	0.03	0.01	0.08	0.08	0.23	5×10^{-4}	260%
16	16	DIAS0001721 (20,774,580)	ALGA0089643 (22,976,914)	Basic	0.02	0.01	0.08	0.09	0.21	1×10^{-4}	171%

% σ^2_{WIN} /Baseline σ^2_A shown for the actual model estimates of σ^2_A regardless of significance (†P>0.1) Win LRT P value significance thresholds are 5.64×10^{-5} at the genome wide and 5.64×10^{-4} at the indicative level.

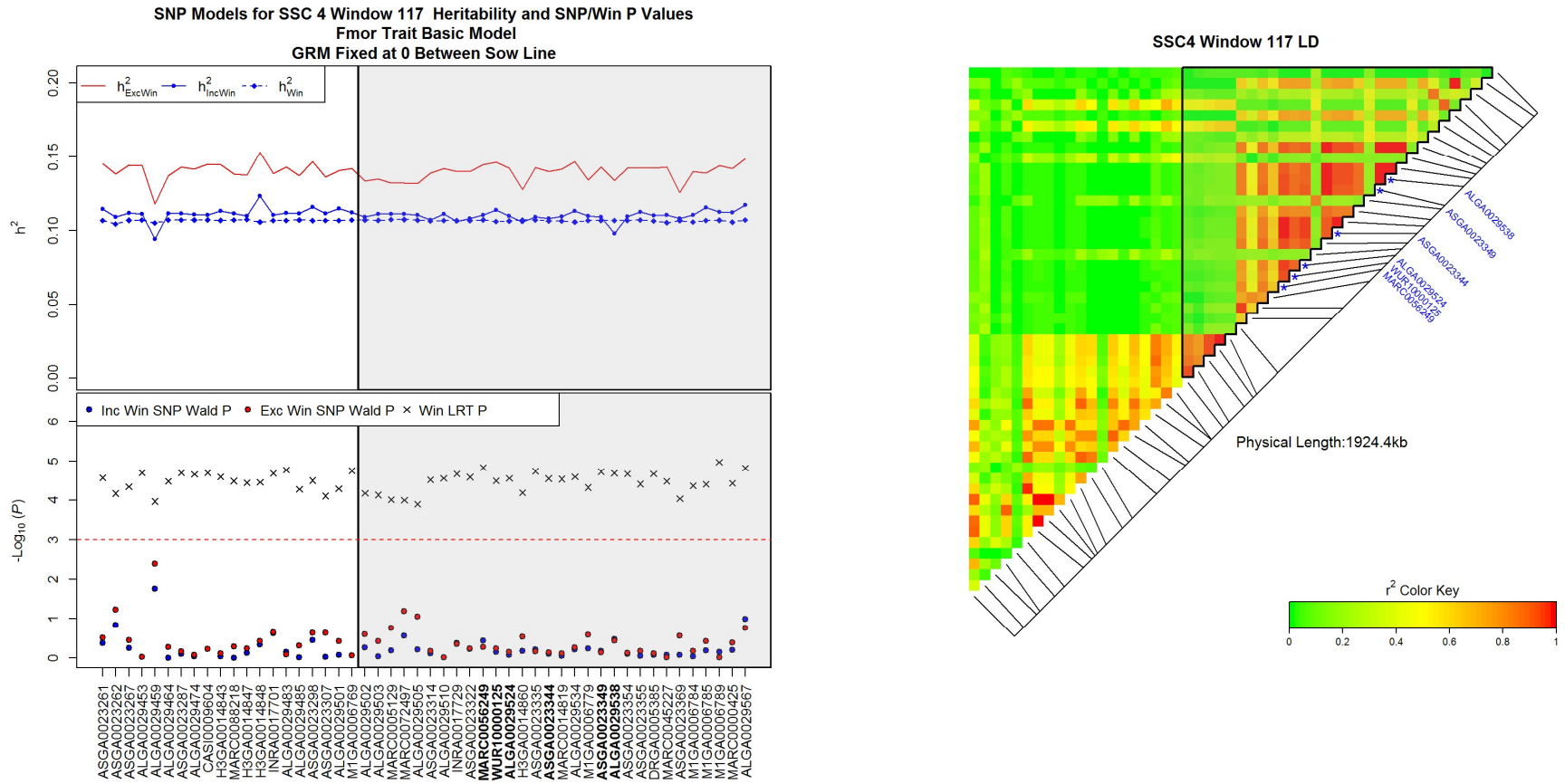

Figure 6.9 – Measured Genotype Analysis for SSC 4 Window 117 and LD Highlighting Association First Published in Boddicker et al. 2012

(Left) Measured Genotype Analysis. Estimates generated with the reduced model (red) and full model (blue). (Top) h^2 estimates (solid line) and window h^2 (dashed line). (Bottom) -$\log_{10}$(P) values shown using the Wald F statistic for SNPs and LRT for window. Bonferroni corrected window wide SNP significance (red dashed line), (Right) LD heatplot of the window. Region identified in association with weight gain and viral load in Boddicker et al. 2012 shown shaded, 6 top SNPS from that study shown bold (left) or labelled (right).

A measured genotype analysis using the RHM data was conducted on all SNPs in the other windows showing a significant effect at the indicative threshold.

Only one candidate SNPs was identified (ASGA0048302) in SSC10 window 46 (Wald P=0.001). It's impact on the heritability estimates was only moderate (in the reduced model), the h^2 reduced from 0.11 (s.e. 0.06) to 0.09 (s.e. 0.06) and in the full model, the h^2_{Win} reduced from 0.18 (s.e. 0.04) to 0.14 (s.e. 0.05)). This SNP was therefore taken forward for further characterisation and validation.

6.03.6 Validation and Effect Size Calculation

SNPs significant at the window wide threshold in the previous measured genotype analysis were characterised further using all traits using the full population GRM. SNPs identified on the same chromosome (both within this analysis including the genome wide level SNPs identified in the previous chapter) were checked for independent effects and LD was explored. 6 SNPs indicated effects in this analysis across the *Still, Dead, Alive* and *Fmor* analyses at the window wide significance threshold, these were on SSC7, SSC9 and SSC10. In two windows (SSC9 window 90 and SSC10 window 57) two SNPs in the same window are shown to have a significant (P<0.001) effect on *Still* and *Dead* trait variance respectively. An additional SNP on SSC10 in window 46 was shown to have a significant effect on *Fmor* trait variance. One further SNP (MARC0095146) in SSC7 Window 91 was shown to have a significant effect on *Alive* trait variance.

The baseline variance components estimates were generated using the data for the subsequent measured genotype analysis is shown in Table A.13. This differs from the RHM data in that it contains repeated records, is run on all animals with genetic data in epidemic phase, includes unmapped SNPs (for QC see section 'FASTA & Measured Genotype QC' on pg.161) and maintains all relationships in the GRM. Only marginal differences are seen in the estimates as compared to the corresponding estimates in Table 6.1.

SNP MARC0095146 identified on SSC7 in the Alive trait analysis was fitted as a fixed effect for all traits separately, using all the available phenotype data, including repeated records and fitting a permanent environmental effect. SNP model estimates are shown in Table 7 including statistics from the mixed model. Only the additive model indicted a significant SNP effect using a T-test so only this model is presented.

Table 6.6 – Variance Components and SNP Effect for MARC0095146 Single SNP Model

Trait	Variance Component Estimates				Additive SNP Effect			
	SNP Wald P	σ^2_A (SE)	σ^2_P (SE)	h^2 (SE)	Additive Model P-Value	α Estimate (SE)	σ^2_{SNP} ($2pq\alpha^2$)	%VA
Mum	0.122	0.02 (0.03)	0.56 (0.03)	0.04 (0.05)	0.05	-0.10 (0.05)	4×10^{-3}	11.18%
Still	0.791	0.09 (0.03)	0.43 (0.02)	0.2 (0.06)	0.25	-0.04 (0.05)	7×10^{-4}	0.84%
Dead	0.253	0.04 (0.03)	0.65 (0.03)	0.06 (0.05)	0.05	-0.11 (0.05)	4×10^{-3}	9.08%
Alive	0.009	0.61 (0.57)	10.7 (0.53)	0.06 (0.05)	7×10^{-3}	0.65 (0.23)	0.16	17.53%
Tof	0.493	0.97 (0.52)	8.8 (0.45)	0.11 (0.06)	0.25	0.20 (0.21)	0.02	1.65%
Fmor	0.105	0.01 (5×10^{-3})	0.09 (5×10^{-3})	0.09 (0.05)	0.02	-0.05 (0.02)	1×10^{-3}	9.84%

SNP effects calculated using measured genotype for SNPs identified as significant at the genome wide level, fitted to all traits. Model Includes basic model terms and repeated measures, with permanent environmental effect. Also includes Wald P statistics for SNP interactions with potential sources of confounding, estimated in separate models.

While MARC0095146 can only be demonstrated to have a significant effect on variance of the *Alive* trait, a significant difference is seen in the ASE ($P<0.05$) for the *Mum, Dead* and *Fmor* traits. Also tested in two separate analyses were the SNP with SNP × farm interaction and SNP with SNP × sow line interaction to explore confounding of the effect with population structure. No SNP interactions with population model terms (sow line or farm) are significant and the MARC0095146 retains the same level of significance.

In the *Still* trait analysis 2 SNPs (ASGA0105527 and DIAS0004436) in SSC 9 Window 90 are shown to have a window-wide significant effect on trait variance. As can be seen in Figure 6.10, the two adjacent SNPs have indeed a higher level of LD with each other than with the remaining SNPs in the same window.

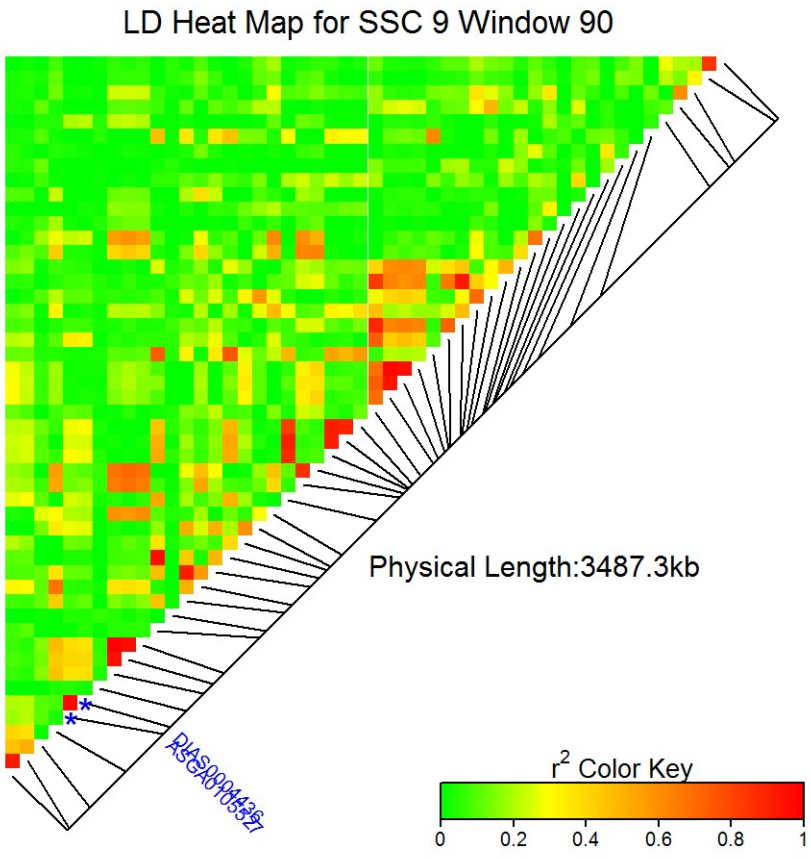

Figure 6.10 – LD Heat Map for SSC9 Window 90

SSC10 Window 57 heat map of r^2, two SNPs significant at the corrected threshold shown labelled (actual r^2 value 0.97 for labelled SNPs). All 50 SNPs in the window plotted.

To investigate whether one of the SNPS indicates a specific effect on the genomic trait variance, genotypes for both SNPs were fitted together as fixed effects in the model. Based on these results DIAS0004436 was subsequently dropped from the analysis.

The remaining SNP, ASGA0105527 has already been indicated at the chromosome wide level using the FASTA method in Chapter 5, (Figure 5.2). This window is only ~7Mbps from the MARC0016053 SNP showing a genome wide association identified using the GRAMMAR method (Figure 5.11). To check whether ASGA0105527 effect is independent of the previously demonstrated association for MARC0016053, an additional analysis was conducted, in which the genotypes of the both SNPs (ASGA0105527 and MARC0016053) and their interactions were included in a validation step. The SNP×SNP interaction was not significant across all traits and was dropped. The Wald P values and variance component estimates fitting ASGA0105527 and MARC0016053 is shown in across traits using the basic model.

Table 6.7 – Wald P values and Variance Component Estimates Fitting ASGA0105527 and MARC0016053 in Full Phenotype Data SNP Models

Trait	SNP Wald P		Variance Component Estimates			
	ASGA0105527	MARC0016053	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	h^2 (SE)
Mum	0.588	<0.001	0.03 (0.03)	0.1 (0.07)	0.43 (0.07)	0.05 (0.05)
Still	<0.001	0.236	0.07 (0.03)	2×10^{-7} (1×10^{-8})	0.34 (0.02)	0.17 (0.06)
Dead	0.262	0.001	0.04 (0.03)	0.26 (0.07)	0.34 (0.07)	0.06 (0.05)
Alive	0.890	<.001	0.65 (0.58)	4.43 (1.3)	5.53 (1.16)	0.06 (0.05)
Tof	0.865	0.763	0.96 (0.52)	2.71 (1.13)	5.15 (1.04)	0.11 (0.06)
Fmor	0.421	<.001	8×10^{-3} (5×10^{-3})	0.04 (9×10^{-3})	0.04 (8×10^{-3})	0.09 (0.05)

SNP models fitting two SNPs showing an association, for results pertaining to MARC0016053 identified using the GRAMMAR method see section 5.03.3.

The ASGA0105527 genotype shows a significant association with the *Still* trait but not the other traits. An independent effect for MARC0016053 is seen, as previously demonstrated the traits *Mum, Dead, Alive* and *Fmor.* However, the effect of MARC0016053 on the *Still* trait variance previously indicated (P=0.097 see Table 5.8) no longer indicates an effect in the two SNP model.

The ASGA0105527 *Still* trait SNP effect size was calculated using a measured genotype model. The allele substitution effect (α) was estimated at -0.14, and the SNP accounted for 6.42% of the total additive genetic variance and 24.46% of the SSC9 window 90 variance.

1 SNP showed a significant association with the *Fmor* trait in SSC10 Window 46 (ASGA0048302). This SNP was also fitted for all traits and showed a significant effect on the trait variance for *Mum* (P=0.003) *Dead* (P=0.005) Alive (P=<0.001).

Given the proximity of window 46 to window 57 (~9Mbps), and therefore the 2 SNPS (ASGA0083169 and MARC0020386) found to have a window-wide significant effect on the *Dead* trait, the potential for confounding of these effects was considered. Figure 6.11 shows the LD for the whole region from window 46 to window 57 and the LD in both of the labelled windows.

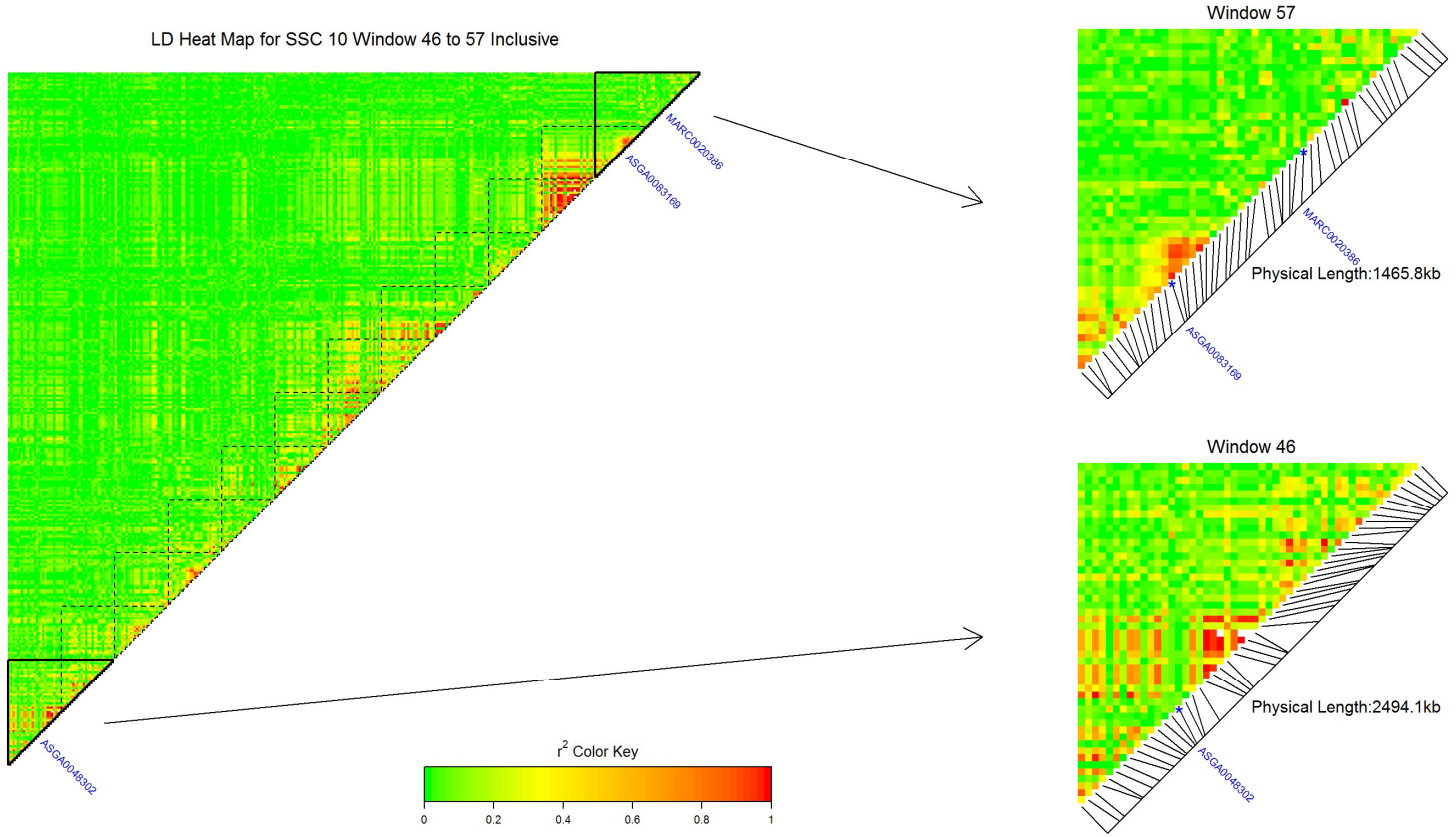

Figure 6.11 – LD (r^2) for SSC10 Windows 47 to 56

(Left) LD (r^2) shown for region covered by SSC10 windows 46 to 57, windows 46 and 57 (solid border) intermediate windows (dashed border). (Right) Individual window LD shown right, three SNPS identified as significant at the window-wide level shown labelled.

Whilst patterns of LD are observed in blocks between the windows, r^2 is low between the SNPs themselves with values between ASGA0083169 and ASGA0048302 of 0.098; MARC0020386 and ASGA0048302 of 0.013; and between MARC0020386 and ASGA0083169 of 0.062.

The independent SNP effects fitting only the individual SNPs MARC0020386, ASGA0083169 and ASGA0048302 in a single SNP measured genotype model are shown in Table 6.8. All three SNPs showed significant effects under the additive model. Specifically, for MARC0020386 significant effects on trait variance are seen for the *Still, Dead* and *Fmor* traits. For ASGA0048302 significant effects on trait variance are seen for the *Mum, Dead*, Alive and *Fmor* traits. Finally, ASGA0083169 has a significant effect on *Mum, Still, Dead, Alive* and *Fmor* trait variance.

Table 6.8 – Variance Components and SNP Effect for MARC0020386, ASGA0048302 and ASGA0083169 in Single SNP models.

Trait	SNP	Variance Component Estimates				Additive SNP Effect			
		SNP Wald P	σ^2_A (SE)	σ^2_P (SE)	h^2 (SE)	Additive Model P	α Estimate (SE)	σ^2_{SNP} ($2pq\alpha^2$)	%VA
Still	MARC0020386	0.015	0.08 (0.03)	0.42 (0.02)	0.19 (0.06)	0.03	-0.21 (0.10)	7×10^{-3}	8.33%
Dead	MARC0020386	0.002	0.05 (0.03)	0.64 (0.03)	0.08 (0.05)	0.03	-0.28 (0.12)	0.01	24.19%
Fmor	MARC0020386	0.023	0.01 (5×10^{-3})	0.09 (5×10^{-3})	0.11 (0.05)	0.11	-0.07 (0.04)	8×10^{-4}	8.47%
Mum	ASGA0048302	0.003	0.03 (0.03)	0.09 (0.08)	0.05 (0.05)	1×10^{-3}	0.19 (0.06)	0.01	59.28%
Dead	ASGA0048302	0.005	0.04 (0.03)	0.26 (0.07)	0.06 (0.05)	2×10^{-3}	0.19 (0.05)	0.01	57.59%
Alive	ASGA0048302	<.001	0.62 (0.56)	4.54 (1.28)	0.06 (0.05)	1×10^{-3}	-0.87 (0.25)	0.28	60.32%
Tof	ASGA0048302	0.121	1.03 (0.52)	2.68 (1.11)	0.12 (0.06)	0.06	0.45 (0.22)	0.07	16.73%
Fmor	ASGA0048302	0.001	0.01 (5×10^{-3})	0.05 (0.01)	0.08 (0.05)	6×10^{-4}	0.08 (0.02)	3×10^{-3}	46.99%
Mum	ASGA0083169	0.011	0.04 (0.03)	0.09 (0.08)	0.07 (0.05)	7×10^{-3}	0.12 (0.04)	7×10^{-3}	19.16%
Still	ASGA0083169	0.015	0.08 (0.03)	2×10^{-7} (1×10^{-8})	0.2 (0.06)	7×10^{-3}	0.12 (0.04)	7×10^{-3}	7.87%
Dead	ASGA0083169	<.001	0.05 (0.03)	0.25 (0.07)	0.08 (0.05)	6×10^{-4}	0.17 (0.047)	0.01	25.24%
Alive	ASGA0083169	0.012	0.99 (0.59)	4.28 (1.29)	0.09 (0.05)	1×10^{-3}	-0.65 (0.20)	0.18	16.58%
Fmor	ASGA0083169	0.002	0.01 (5×10^{-3})	0.04 (0.01)	0.1 (0.05)	4×10^{-3}	0.05 (0.02)	1×10^{-3}	10.06%

SNP effects calculated using measured genotype for SNPs identified as significant at the genome wide level, fitted to all traits. Model Includes basic terms and repeated measures, with permanent environmental. Also includes Wald P statistics for SNP interactions with potential sources of confounding, estimated in separate models,.

To test for confounding of the three SNP effects, the measured genotype analysis was repeated for all reproductive traits with all three SNP genotypes included in the model together with all possible 2 and three-way interactions. The three-way interaction and the ASGA0048302 × MARC0020386 interaction were not significant in any of the models. However, some of the 2-way interactions were found significant (ASGA0048302 × ASGA0083169 acting on *Still*, *Dead*, *Alive* and *Fmor* trait variance, and MARC0020386 and ASGA0083169 acting on *Tof* trait variance), thus indicating epistatic effects.

6.04 Discussion

In total 17 windows were significant above the indicative threshold across the RHM analyses presented for 5 traits. One window on SSC4 was significant at the genome wide level covering 138Mbp to 140Mbp. Using the measured genotype analysis on SNPs contained in these windows, six SNPs showed significance at a window-wide level. Exploring these SNPs further, 5 independently acting SNP effects were detected, 2 of which also indicated a potential epistatic effect.

A genome wide significant association with *Fmor* has been shown on SSC4 between 138,2911,968 and 140,216,348 bps. This location coincides with a well characterised association previously reported on this chromosome with viral load and weight gain of growing pigs infected with a virulent strain of PRRSV (Boddicker *et al.,* 2012). Whilst a direct association with the WUR10000125 genotype could not be shown, different patterns of LD are reported in this study population to those reported in Boddicker et al. 2012. These results indicate that genetic variation in the same region as that reported in Boddicker et al. 2012, associated with viral load and weight gain, may also translate to effects in terms of foetal mortality. Several studies using single SNP methods have investigated variance associated with the WUR1000125 genotype in other populations (Serão *et al.,* 2016; Yang *et al.,* 2016) which failed to show associations in this region. Given that these methods rely on

LD between the causative mutation and a single SNP, linkage is not maintained across study populations, and therefore an effect with the same locus may not be found.

Four individual SNP associations were identified using measured genotype analysis on windows significant above the indicative threshold.

ASGA0105527, also indicated at the chromosome wide level using the FASTA method (see Chapter 5) was identified on SSC9 in window 90. This showed a significant ($P<0.001$) effect on *Still* trait variance. The allele substitution effect (α) was estimated at -0.14 and SNP variance (σ^2_{SNP}) calculated at 6.42% of the total additive genetic variance σ^2_A and 24.46% of the σ^2_{Win}. No effect in this SNP was observed for any of the other disease indicator traits.

MARC0095146 was identified using the *Alive* trait in window 91 on SSC7. This SNP explained 17.53% of the total *Alive* trait σ^2_A. Whilst this SNP could not be demonstrated to have a significant effect on the variance of other traits, significant differences in predicted genotype means were indicated for the *Mum, Dead* and *Fmor* traits.

Several windows indicated significance on SSC10, using the *Dead* and *Fmor* traits. In window 57 MARC0020386 showed evidence of an association with the *Still, Dead* and *Fmor* trait variance ($P<0.05$). Significant differences were also seen for the ASE in the *Still* and *Dead* traits. Also on SSC10 two SNPs, ASGA0048302 and MARC0020386, showed evidence of an association in window 46, with low LD found between the SNPs. Both the SNPS and the SNP×SNP interaction showed a significant effect on trait variance for the *Dead* and *Fmor* traits, with significant effects also indicated using other traits in either the SNPs or the interaction term. This could suggest a more complicated genetic architecture at this location.

These results show some correlation with the previous single SNP GWA analysis. In addition to ASGA0105527 for which a putative association is shown, also indicated in the FASTA *Still* trait Epidemic Phase Joint Analysis, several windows show cross over with

several SNPs significant in Chapter 5. SSC 5 Window 54 significant in the *Still* trait analysis contains SNP MARC0044148 indicated at the chromosome level in the *Wean* trait analysis, though no direct SNP effect could be observed using measured genotype analysis. Also, SSC 8 Window 23 indicated in the Alive trait analysis is 5Mbp from MARC0098261 indicated in the Basic Model *Wean* trait GRAMMAR analysis.

In addition to the overlap with the previously published associations on SSC4, a few other windows significant at the indicative level are close to some of those reported in previous PRRSV genomic studies. Identified in *Still* trait analysis SSC5 Win 54 is ~2 Mbps from a single SNP association published in (Yang *et al.,* 2016), additionally window 80 on SSC14 is ~4 Mbps downstream from a single SNP association in the same publication. The two windows identified on SSC 10, windows 46 and 57 are ~7 Mpbs upstream and ~1 Mbps downstream of a region identified on that chromosome in association with weight gain (Boddicker *et al.,* 2014a).

The estimate of window variance as a percentage of additive genetic variance was generally high, in some cases >100%, implying that the genomic region in consideration absorbed the majority of the genetic variance. This was particularly the case, as described when the alternative models Alt1 and Alt2 were fitted. Lower regional than whole genomic variances were however observed for the *Still* trait. Coincidentally, for this trait variance estimates using genomic relationships were also more similar to the estimates using the pedigree based method for farm 1, which form the majority of data in the joint farm analyses. In line with previous "missing heritability" observations (Chapter 4), one may hypothesise that the genomic estimates of the total additive genetic variance are downward biased for the majority of traits, and this may partly explain why regional variance estimates were relatively large.

The three models explored in previous chapters were explored in the RHM analysis. QQ plots suggest these models differed in their ability to detect window effects. For the *Mum*

and *Dead* traits, some evidence is seen in improved ability to detect significant windows using alternative models, for other traits the reverse is true. Generally, differences in the significance associated with each peak under the different models suggests slight changes in emphasis. Some peaks are robust to the effect of fitting the alternative models, in some cases the order by rank of the significant windows changes and for some windows different window effects are indicated by the Alt.1 and Alt.2 models. Whilst no systematic effect can be observed this change in emphasis is interesting.

Considerable reductions in additive genetic variance are seen fitting the alternative models on the baseline estimates seen in Table 6.1. In the RHM analysis where windows are indicted in several models the additive genetic variance is relatively consistent between models, this leads to very large values for window variance as a proportion of total baseline genetic variance estimates.

Six associations have been presented, one of which is shown above genome-wide significance threshold, demonstrating for the first time the genetic potential of this well studied region in mitigating foetal mortality during PRRSV outbreaks. Whilst findings at the indicative level may lack conclusive evidence of an effect, the five SNPs identified in these windows show a broad effect across the traits analysed, providing additional evidence supporting a putative QTL effect in these regions. Additional evidence is seen for the effect on SSC 9 in the single SNP GWA analysis, and for the effect on SSC10 in independent analyses.

Chapter 7. General Discussion

7.01 Aims and Outcomes

7.01.1 Thesis Objectives

The aims of this thesis were to explore the genetic architecture underlying reproductive outcomes to PRRS:

- To consider the factors affecting non-genetic effects contributing to reproductive trait variance in PRRS challenged data, such that models may be devised to explore additive genetic effects.
- To quantify additive genetic variance and heritability, using pedigree information (where available) and 60K SNP genotypes.
- To utilise single SNP association methods to explore the genetic architecture of reproductive trait variance across the pig genome
- To use regional association methods, to look at the reproductive effects in wider genomic windows

7.01.2 Key Research Outcomes

- Evidence was presented suggesting that pre-exposure to PRRSV could reverse a pattern of higher PRRS-associated reproductive losses in later parities in naïve animals.
- Additive genetic variance and heritability were estimated for reproductive traits in the absence and during PRRS outbreaks, and it was shown that accounting for epidemiological trends tends to reduce these estimates.
- A more in-depth analysis of the genomic relationship matrix (GRM) structure showed that between sub-populations relationships could be affecting the ability to accurately estimate the additive genetic variance.

- Two Single SNP GWAS methodologies (FASTA and GRAMMAR) were employed and 5 SNPs with significant associations with reproductive traits a genome-wide level were detected using GRAMMAR:
    - SNP ALGA0087207 on SSC15 showed a genome-wide significant effect on the variance of the number of Stillborn piglets per litter. This SNP is 4 Mbp equidistant between two SNPs previously demonstrated to be associated with foetal mortality when this trait is considered as a property of the dam
    - The four additional associations that were significant at the genome wide level were with SNPs on SSC3, SSC9, SSC12 and SSC17.
- Regional heritability mapping analyses were conducted on survival and mortality traits, with an in-depth analysis of how each SNP impacted each genomic window, and total additive, genetic variance. Within windows significant above the indicative threshold, five putative SNP effects, significant at a window wide level, were identified.
    - A genome-wide significant window effect was identified on farrowing mortality in a region on SSC4, previously shown to be associated with PRRS pathogen burden and weight gain during infection.
    - Evidence of an effect on reproductive trait variance was found for a region on SSC10, 1 Mbp from a locus previously demonstrated to be associated with weight gain during PRRSV infection. In this window 2 independent SNP effects were identified which may indicate an epistatic effect.
    - In windows significant at the indicative level, additional putative SNP associations were found with loci on SSC7, SSC9 and a further association on SSC10, significant at the window wide level.

7.02 Impact

As discussed in discussed in Johnston *et al.,* (2003), whilst experimental trials such as controlled pathogen challenges offer considerable benefits to the scientist in terms of the ability to control experimental design it is often unclear how findings in these artificial situations, may translate to a production setting. This difference may be considered analogous to the realisation of traits in a field setting as compared to the selection under idealised conditions in a nucleus herd. As highlighted in Zumbach *et al.,* (2007) the genetic correlation between nucleus herds and commercial settings is of considerable importance to the animal breeders along with performance of elite genotypes in different environmental conditions. These differences in performance by environment are considered under the genetic (or genotype) by environment interaction (G×E), (Falconer & Mackay, 1996; Lynch & Walsh, 1998). This issue is especially pertinent in an infectious disease situation where the environmental force of infection is dynamic. As such effects observed under experimental conditions (such as has been observed on SSC4 associated with the WUR10000125 genotype) should be reproducible in a field setting to validate that any benefit can be realised under these dynamic and more noisy conditions. Analyses on the economic impact of PRRS suggests reproductive losses form the greater component of costs associated with the disease compared to those associated with reduction in growth (Holtkamp *et al.,* 2013). As such improving the reproductive outcomes during PRRS outbreaks would confer direct financial rewards in animal production systems.

As observed in (Lunney *et al.,* 2016) previous analyses have been unable to show an effect on reproductive performance for the WUR10000125 genotype on SSC4 (Serão *et al.,* 2014; Yang *et al.,* 2016). The results from this thesis indicate for the first time that the previously documented region on SSC4 associated with weight gain and viral load, could also impact reproductive outcomes in swine production systems Whilst a direct association with WUR10000125 could not be demonstrated using RHM the region containing both the

WUR10000125 genotype and the gene which is taken to be the causative mutation show a genome wide effect on farrowing mortality. In addition, two further loci on SSC10 and SSC15, where PRRS related effects have been also been detected in independent datasets, merit further consideration. The use of more than one locus in the genetic improvement of livestock for disease traits is desirable to reduce the likelihood of a specific evolutionary response in the pathogen to overcome the mechanism by which disease is being evaded (Bishop & MacKenzie, 2003). In addition, several novel regions are presented which could assist in the development of a broader genomic improvement programme.

An observation is also made which could impact management strategies, suggesting pre-exposure to the virus may mitigate higher losses in later parities in naïve animals.

In addition to the key research outcomes (with regards to the stated objectives) the results suggest that the partitioning method mainly based on performance data (i.e. independent of formal diagnosis or confirmed aetiology) presented in this thesis could be successfully employed to identify periods that inflict reproductive stress on a large proportion of animals in the population. Using this method, periods of elevated levels of reproductive failure could be identified, and the positive impact of infection on trait heritability could be established This suggests that the trait trend partitioning method could be extended to a wider spectrum of sources of stress to consider a wider genetic architecture of robustness to reproductive stress.

7.02.1 Traits Used and Pathogenesis of Disease

One of the main differences between this and other similar studies on PRRS reproductive failure (i.e. Lewis et al., (2009a, 2009b) and Serao et al., (2014)) is the inclusion of *Tof* as a fixed covariate. This changes the emphasis of the analysis presented which accounts for a simple proportional relationship between the trait under evaluation and litter size. Results indicate that this may be beneficial in terms of identifying additive genetic variance

attributable to the pathogenesis of disease. In the non-epidemic phase some of the traits show a low but estimable heritability without *Tof* fitted (i.e. *Alive*) at ~0.075 (±0.005) across analyses which are considerably reduced when *Tof* was included. This contrast in additive genetic variance was used to explore the potential underlying genetic liabilities associated with PRRS related mortality and survival.

In this analysis as with others (Lewis *et al.,* 2009b; Serão *et al.,* 2014) reproductive performance was considered a trait of the sow, rather than trait of the individual piglets in the litter, as used e.g. in Yang *et al.,* (2016). The relative merits of these approaches may depend on the pathogenesis of disease in the reproducing sow. As discussed in the introduction the mechanism of piglet death following PRRS outbreaks has not been fully explained. Whilst the explanations provided by some authors suggest that the biology of sow is the main factor in determining foetal mortality (Karniychuk & Nauwynck, 2013) others have indicated that the outcome maybe down to active infection of the individual piglets, and are (at least in part) determined by the biology of the piglet (Ladinig *et al.,* 2015a; Lunney *et al.,* 2016). Whilst these experiments lacked the piglet genotypes, animal infection statuses or measures of viraemia to tackle this issue head on, it's potential impact on the estimation of additive genetic variance is considered. The underlying source of the genetic liabilities would determine the correct experimental design to use in estimating reproductive mortality during PRRS outbreaks: If sow biology is the sole factor in determining piglet mortality then one might expect the sow model to better capture the underlying genetic liability of reproductive performance. Vice versa if the piglet biology is the sole factor in determining piglet mortality then one might expect the piglet model to better capture the underlying genetic liability of reproductive performance. If either of these hypotheses were absolute (i.e. under total sow or piglet control), given the amount of genetic material shared by the piglet and sow is approximately ½ one might expect the use of the 'incorrect' model to reduce the estimates of additive genetic variance by this amount. However, given that the virus must pass through

the sow to reach the pig some aspects of sow resistance (when defined as ability to reduce virus load) would undoubtedly play a role in the liability associated with piglet death (i.e. exposure), however tolerance may not. Whilst these are fascinating questions in terms of the reproductive performance of sows to PRRS outbreaks they were beyond the scope of this study.

7.02.2 Statistical Models Used

One of the issues of retrospective farm analyses and field studies remains the difficulty in controlling confounding variables and the inability to adjust for these in statistical analyses (Johnston *et al.,* 2003). Yet to understand the genetic effects on an outcome it is necessary to account for other non-genetic factors involved in trait variance. In my analyses, several models were used. The basic model (Basic) was based on backward elimination stepwise regression of animal characteristics in the available data. In the alternative models the between epidemic effect (Alt.1) and a dynamic within epidemic effect based on a trait trend (Alt.2) was additionally considered.

The LRT results indicated improvements in the model fit when accounting for individual epidemic effects and epidemic trend using linear mixed model analysis. However, it remains difficult to conclude which of the three models performed best in terms of estimating genetic parameters. This is most likely the result of confounded variables and the differences in population structure between the farms. Farm 1 had considerable diversity in terms of breed and (sow) line not seen on farm 2. On farm 1 and in the joint analysis, when the individual epidemic effect was considered (Alt.1) a reduction in additive genetic variance was seen. It is possible that several factors (e.g. viral strain, susceptibility and genetic by environment interactions) were confounded in fitting Epidemic ID. Given only one epidemic occurred on Farm 1 the Alt.1 model was not applicable to this analysis. On Farm 1 and in the joint farm analysis, fitting Epidemic ID and dynamic trait trend (Alt.2) absorbed a large proportion of the additive genetic variance, resulting in very low heritability estimates. On farm 2 models

fitting the dynamic trend (Alt.2) absorbed only environmental variance resulting in higher heritability estimates.

The Alt.2 model was established to investigate the effect of the individual whilst accounting for a contemporary effect of animals expressing the phenotype at a similar time. One may expect that a contemporary group effect has a greater genetic component in more genetically diverse populations. For the farm 1 and joint farm analysis a reduction in the additive genetic variance is seen using alternative model 2. A downward bias to heritability estimates has been shown for resilience type traits if cohort estimates of disease burden are confounded with resistance (Doeschl-Wilson *et al.,* 2012). Fitting the trait trend based on these analyses, was suggested by our research group for use in Serão *et al.,* (2014), who found improved ability to account for an environmental contemporary effect compared to the conventional Herd-Year-Season effect. Those findings are similar to those reported here in farm 2 analyses.

It was thus not possible to ascertain the "best model" for estimating genetic effects. This may depend on population structure and the complex genetic interactions which may occur between genetically diverse animals. The observed reduction of the genetic estimate may relate to a specific pattern of resistance and the resulting variance component estimates may still hold interesting information with regards to the genetics of response in the host. As such these models were further explored using association methods.

With regards to the association analyses there was some evidence of sensitivity to the model used in the farm 1 and joint farm populations. In the joint analysis, several loci identified (e.g. ALGA0121571 using GRAMMAR and SSC1 window 151 and SSC10 window 57 using RHM) suggest improved ability to detect these effects using the Alt.2 model where the levels of significance reached were greater. This could indicate an improved ability to detect genetic effects in models using the Alt 2 models. All SNP effects detected using the Alt.2

model subsequently proved significant when as a fixed effect with the Basic and Alt.1 models.

Further studies would need to be undertaken to look at the effects of fitting this type of cohort mean in genetic analyses. Simulation methods may prove useful for this purpose. It would also be interesting to see if associations showing sensitivity to the Alt.2 in the association methods indicate different effects with regards to specific mechanisms of resistance or tolerance.

7.02.3 Management Strategies

Results from general linear mixed models could indicate a systematic pattern of higher losses in higher parities in naïve animals; and that this pattern is reversed in animals with previous exposure. These findings differ from those reported in Lewis *et al.,* (2009a). In the present analysis, a slight modification of the partitioning method was used compared to the one in the previous study. This considered epidemic-reproductive failure associated with PRRS outbreaks, resulting in the identification of three epidemics, as compared to the two PRRS outbreaks used in Lewis *et al.,* 2009a. Given an ELISA negative result following an outbreak depends on the seroreversion of animals after clearing the virus (taking several months); ELISA alone may thus not provide the resolution to distinguish between epidemics less than a few months apart. In this analysis two separate epidemics, approx. three months apart were analysed, considered one outbreak in the previous analysis. In Figure 2.10 animals in the first epidemic show an increase in the number of mummified piglets with increasing parity. In the third epidemic animals show a decrease in the number of mummified piglets with increasing parity. Animals in the second epidemic show a mixture of the two patterns with early parity animals (≤4) showing a pattern similar to that in the first epidemic with later parity animals (≥6) showing a considerable reduction. It is possible that the later parity older animals were present on the farm at the time of the first epidemic. Whilst the risk of pre-exposure was considered as a possibility by Lewis *et al.,* (2009a), to test this, older animals in the second

outbreak were removed. It is suggested in this analysis that the two epidemics considered in that outbreak combined two different patterns of pre-exposure resulting in the non-systemic parity effect observed in the previous study. Whilst the differences in the model are acknowledged (in the correction for *Tof*), very similar parity profiles were obtained using the basic model as those presented in Lewis et al., (2009a), it was only when considering three distinct epidemics that the results suggest a more systematic epidemic specific parity effect.

Whilst this explanation fits observations of early parity (≤4) (presumed naïve), sows on Farm 2) a slight inconsistency is observed in parities (≥5). A non-monotonic profile is seen which may indicate some confounding of exposure by age group in these cohorts.

Given a lack of individual disease statuses, and the possibility of confounders (e.g. different co-infections in different epidemics or different circulating viral strains) further work would be required to formally establish a link between previous exposure and a change in parity effect.

I think it unlikely that confounding is responsible for different systematic parity effects observed in these animals and that pre-exposure provides an adequate explanation for the observation. It would be interesting to see if the same reversal of parity effect is seen in systems where vaccination is used.

This pattern would suggest, that a practice of challenging animals during acclimation with PRRSV (for example, by autogenous vaccination of gilts with pooled sera from the older sows as currently used on some farms), could mitigate against considerably higher losses where reproducing sows become infected in higher parities. Whilst this practice is performed on some farms where eradication is not an objective these findings may assist in the cost/benefit decision making for the objectives of PRRS control. Farms experiencing PRRS epidemics at a low frequency with greater numbers of older PRRS naïve animals, are likely to see greater impacts when outbreaks occur.

7.02.4 Resistance, Tolerance, Resilience, Robustness

Resistance is generally accepted as the ability of the host organism to limit pathogen burden, whereas resilience is performance despite pathogen burden (Albers *et al.,* 1987) or, mathematically; the slope of the regression of performance on disease burden (Simms & Triplett, 1994; Bishop, 2012). The broader term resilience to disease can be used to describe performance in the presence of disease challenge and the broader term, robustness, used to describe the ability to perform despite external stressors affecting performance (Knap, 2005; Bishop & Woolliams, 2014).

The comparisons drawn in this study consider the performance of animals in the presence and absence of episodes of high rates of reproductive failure associated with PRRS outbreaks. As such, based on purely this definition, the resilience of the sows to PRRS is considered. However, given a lack of information on disease burden and the open question on the pathogenic mechanism of reproductive failure, defining response in terms of resistance and tolerance are beyond the scope of this study. Where such issues cannot be resolved the broader term of resilience is more applicable. It could be argued that to demonstrate a true resilience type effect, the favourable SNP effects should be shown to have no negative effects on reproductive performance in non-epidemic data. Given that no effect (negative or positive) on performance could be demonstrated in non-epidemic phase these SNPs could still be considered as showing a resilience-type effect. The effects of SNPs in non-epidemic phase should be fully quantified prior to use in a breeding programme.

There are a number of causes of reproductive failure in sows including viral infections (pseudorabies virus (PRV), influenza, porcine circovirus type 2 (PCV2) and parvovirus), bacterial infections (leptospirosis) and a range of non-infectious aetiologies (e.g. Mycotoxins, body temperature, age) (Rueff, 2000; Pittman, 2008). Whilst robustness across sources of stress is considered in the literature (Knap, 2005; Archibald *et al.,* 2008), there are no studies on the genetic merit of pigs during periods of stress of different causes.

Using the threshold-threshold method, not only identified periods coinciding with known PRRS confirmed outbreaks but also a period that resulted in reduced reproductive performance of sows despite negative diagnostic results were identified. Whilst it is not possible to categorically preclude PRRS as a potential cause of the reproductive failure in epidemic 5 it is unlikely given the high specificity of the diagnostic test used on the farm. Using the combined periods of "stress" (periods of elevated estimates of *Dead* trait trend) on farm 2, a significant heritable component for *Still* and *Alive* traits are reported at 0.18 and 0.16 respectively.

The partitioning method could easily be used to scan animal production databases by farm to identify periods of increased reproductive stress. Given that pedigree and/or genomic information is in many cases already available for these animals, periods could be combined to investigate the genetic merit of animals across periods of reproductive stress. Studies could consider not only the performance across epidemics but using multivariate analyses that consider the genetic structure between periods and traits. Whilst differences may be expected to be observed in the underlying genetic liability for robustness across aetiologies, common genetic signals may be identified in the GRAMMAR analysis.

7.02.5 Associations

In the joint farm analysis using the GRAMMAR method five independent SNP effects were identified, significant ($P<0.05$) at the genome-wide level:

- ALGA0121571 on SSC 3 (24,409,351 bp[2])
- MARC0016053 on SSC 9 (135,725,781 bp[2])
- ASGA0054360 on SSC 12 (36,158,023 bp[2])
- ALGA0087207 on SSC 15 (138,593,445 bp[2])
- MARC0016887 on SSC 17 (25,600,780 bp[2])

[2] *Sus scrofa* build 10.2 genome coordinates

Using RHM one window showed a genome-wide significant effect on SSC4 (138,291,968-140,216,348 bp). Testing SNPs within windows showing an effect at the window wide level five putative SNP associations were identified.

- MARC0095146 on SSC 7 (111,370,919 bp^2)
- DIAS0004436 on SSC 9 (127,817,182 bp^2)
- MARC0020386 on SSC 10 (69,899,496 bp^2)
- ASGA0083169 on SSC 10 (69,395,371 bp^2)
- ASGA0048302 on SSC 10 (58,071,987 bp^2)

It is acknowledged that the SNP effects calculated are possibly overestimates, as a result of the "Beavis effect" (Beavis, 1998), the results would suggest that these SNPs make considerable contributions to variance in the reproductive outcomes. Additionally, as the full GRM was used in the measured genotype analysis (as compared to the GRM fixed at 0 between populations used in RHM), the baseline additive genetic variance may be under-estimated. This would lead to upward bias in the percentage SNP variance of additive genetic variance.

A comparison of these SNP loci to previously published SNP associations are summarised in a karyotype ideogram (Figure 5.1).

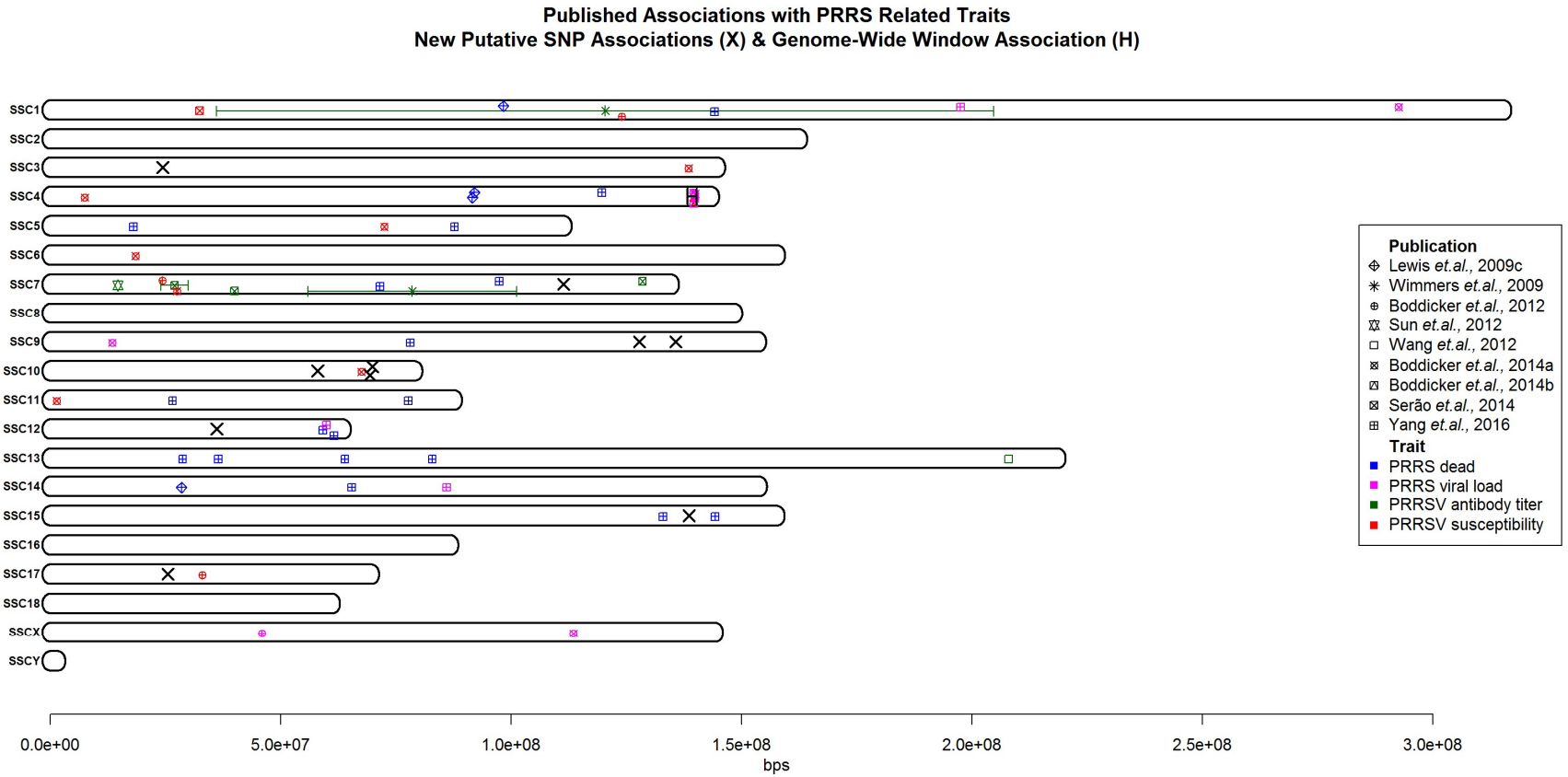

Figure 7.1 – Ideogram Karyotype of the Pig Genome Showing Published Associations for PRRS Traits Including Associations From This Thesis

Data are collected from the pig QTL database (Hu *et al.,* 2005) (see also https://www.animalgenome.org/cgi-bin/QTLdb/index) with information included from Lewis, 2008; Lewis *et al.,* 2009; Serão *et al.,* 2014; Yang *et al.,* 2016. Traits are as reported in the database or taken directly from studies. See Text for further details. The indistinguishable points towards the end of chromosome 4 are a group of papers providing further analysis of the QTL first published for this region in (Boddicker *et al.,* 2012). This plot includes all published associations regardless of the significance threshold employed.

SSC4

A large body of work on the genetics of host responses to PRRSV infection has focussed on the major QTL on SSC4 identified in (Boddicker *et al.,* 2012) which is well documented and has a plausible functional genomic basis through the function of the GBP5 gene family. Given that genetic variation in this region is associated with both increased resistance and tolerance to PRRS, it is puzzling as to why other studies have failed to find an effect on piglet mortality in this region. Results from this analysis suggest that an effect on piglet mortality maybe occur as a result of genetic variation in this region. An investigation of LD in the region show different patterns of LD than those reported by the PHGC. It is therefore possible that SNP WUR10000125 and the causative mutation are not in phase throughout all pig populations, and may have limited the ability of studies on other populations to detect an effect. Further research could investigate genomic selection methods to consider the sum of marker effects in this region and how well these correlate to the regional GRM based estimates of σ^2_{WIN}. This would help elucidate further how to use this region in selection methods where a direct effect with WUR10000125 cannot be detected. Economic analyses estimate foetal mortality as the greatest proportion of losses through PRRS outbreaks, associated with pig production. This research suggests for the first time that that genetic variation in the WUR100000125 region could impact this economically important trait.

In addition to the region on SSC4 two further regions of interest are identified in association with reproductive traits during PRRS epidemics, close to regions of the genome implicated in PRRS response in other studies.

SSC10

Two window associations were indicated using RHM on SSC10, within which three individual SNPs showed a significant effect on trait variance. In SSC 10 Window 57 SNPs MARC0020386 and ASGA0083169 showed significant effects on *Dead* trait variance at a window-wide level. Significant under an additive model SNP variance was estimated at

24.19% and 25.24% of total *Dead* trait variance for MARC0020386 and ASGA0083169, respectively. Additionally, a potential epistatic effect was observed between the SNPs. Effects were also observed for these SNPs for both the *Still* and *Fmor* traits. This association was approx. 1Mbp from a 1Mbp region indicated in Boddicker *et al.,* 2014 ranking 4th in terms of the % marker variance for weight gain in the region as a proportion of the total marker variance. GATA3 was identified as a potential candidate gene in the region lying between the two SSC10 SNPs. The human ortholog of GATA3 has functions in T cell differentiation and promotion of interleukin secretion (Yagi *et al.,* 2011). Moreover, window SSC10 window 46 (11 Mbp upstream) showed an effect on *Fmor* trait variance with ASGA0048302 indicating a window wide effect. A significant allele substitution effect was also observed with a calculated SNP variance representing 10.06% of total genetic variance in the *Fmor* trait.

SSC15

An association with the *Still* trait was demonstrated at the genome wide level for ALGA0087207 on SSC15 using GRAMMAR analysis, this SNP was also indicated in the FASTA analysis at the chromosome level. Using the additive SNP effect model, SNP variance explained 1.79% of the total *Still* trait σ^2_A. Significant effects were seen on the variance of all reproductive disease indicator traits. This SNP lies ~4 Mbp, equidistant, between two associations with foetal mortality reported by Yang *et al.,* (2016) .

SSC3

There was further strong evidence of an effect in these data associated with loci on SSC3 and SSC9 using both single SNP and RHM methodologies. On SSC3 SNP ALGA0121571 (24.4 Mbp) showed a genome-wide significant effect using GRAMMAR analysis, in association with the *Fmor* trait. An effect was indicated on this chromosome using RHM in the *Alive* trait analysis in windows 22, 23 and 24 (27.7 Mbp – 31.8 Mpb), when the GRMs where not fixed at 0 between population this window reached genome wide significance. SNP variance

for ALGA0121571 was estimated at 38.11% of *Fmor* σ^2_A, with effects indicated on variance for all disease indicator traits.

SSC9

Effects associated with two locations ~8 Mbp apart on SSC9 were detected in the RHM analyses Within window 90 on SSC9 SNP MARC0016053 was significant at the genome-wide level in association with the *Mum* and *Fmor* trait using GRAMMAR. Using RHM ASGA0105527 on SSC9 was identified. This SNP also indicated an effect exceeding the chromosome level significance threshold on Farm 1 using the FASTA methodology. An independent effect was observed on *Still* trait variance for ASGA0105527 whilst an effect was observed on *Mum, Dead, Alive* and *Fmor* trait variance for MARC0016053.

Novel Loci on SSC7, SSC12 and SSC17

Finally, three other SNP associations were detected, which did not appear adjacent to previously reported associations for PRRS related traits. One SNP on SSC7 (MARC0095146) was identified using the RHM method, showing a significant effect on *Alive* trait variance. With a significant allele substitution effect using an additive model, SNP variance was estimated at 17.53% of the total additive genetic variance. Whilst an effect for this SNP was not observed on the variance of other traits, a significant allele substitution effect was indicated with the *Mum, Dead* and *Fmor* traits.

Using GRAMMAR, a single SNP was identified on SSC12 (ASGA0054360) in association with the Alive trait explaining 13.56% of the total additive genetic variance. This SNP indicated effects on the variance of all traits with significant additive allele substitution effects shown.

Also using GRAMMAR, a SNP on SSC17 (MARC0016887) showed a genome wide significant association with the *Still* trait. When fitted as fixed effects this SNPs also showed a significant effect on the variance of all traits. A significant ASE was also observed using an

additive model. Although Boddicker et al. (2012) also reported a SNP effect on SSC17, these SNPs are ~7 Mbp apart. Despite the fact that LD blocks are longer in pigs than humans, these SNP associations are unlikely to be marking the same underlying causal genetic variants.

Summary

The 3 loci identified on SSC 4, 10 and 15 are close to or overlapping associations observed on PRRS related traits in independent datasets. Whilst the region on SSC 4 is well characterised and understood, the regions on SSC10 and SSC15 have not been widely explored, these results suggest that further work could be done in these regions to explore the functional basis of host variance and their suitability for use in selection. The recently, released improved pig reference genome sequence (Sscrofa11.1) and associated annotation expected for release in August 2017 should facilitate such searches for positional candidate genes. However, inspection of the human genes that map to the homologous regions of the human genome, did not reveal any compelling functional candidates to date. Strong evidence of QTL effects was observed on SSC3 and SSC9 in these data which would merit further investigation and validation in independent datasets. Three further novel associations are presented which indicate effects in several of the reproductive traits used.

For many of the SNPs identified, some cross over is indicated in terms of effects in traits other than those in which they were identified. This suggests a degree of genetic correlation between pairs of traits. This could be assessed using bivariate analysis to calculate genetic and phenotypic correlations between traits. This forms part on ongoing future work. This work will consider the genetic correlations between traits in the models presented in epidemic and non-epidemic phase and also the effect fitting *Tof* has on the genetic correlations in non-epidemic phase.

7.03 Implementation

Several loci are presented with putative effects on reproductive performance during PRRS outbreaks, whilst one SNP (WUR10000125) is currently being incorporated into breeding strategies, two others are recommended for investigation. The choice of strategy employed for the incorporation of these loci into breeding goals will define the next steps used in implementation. For use in genomic selection the accuracy of genomic prediction based on these loci would need to be considered. This accuracy can be derived using k-fold validation methods to estimate the predictive potential of these regions. This predictive potential quantifies the benefits that may be realised through selection. If used in marker assisted selection it may be desirable to place greater weight on variants with plausible functional effects. To consider such functional aspects it would be necessary to identify the causal genetic variants and characterise their mode of action. One approach would be to look for evidence of differences in gene expression for genes that map to the regions of interest both in the absence and presence of PRRSV between animals of the different predicted QTL genotypes. If differences in transcripts or transcript levels can be identified the genetic and physiological basis of those differences can be explored.

For the other novel loci presented, their effect would need validating in independent datasets prior to incorporation into breeding goals. This could be done by looking at the accuracy of these regions in genomic prediction on reproductive performance in independent populations.

Whilst the detection of additive genetic variance may suggest a use for disease response traits in genomic selection, in practical terms this may not be the case. To maintain the accuracy of genomic selection a feedback process is established in which the realisation of traits in predicted animals are fed back into the validation set to improve the accuracy of the genomic prediction process. This leads to improvements in the trait mean over time. For most economic traits (such as number born alive) these are continuously expressed over the

time course of an animal's life in a production setting. Where PRRS is non-endemic, responses to disease or PRRSV exposure and challenge are only expressed under the specific condition of an outbreak occurring. As such marker assisted selection for genotypes with a known functional basis may be preferable to genomic selection. Further studies to assess the benefits of genomic selection over MAS is required, which should not only include consideration of prediction accuracy but also data availability and genotyping costs.

It may be more attractive to use marker assisted selection to screen the selection candidates to ensure that they are as hardy as possible to disease challenge. This is the strategy successfully employed for the genetic improvement of Salmon to IPN and sheep to scrapie as discussed in the introduction. This judgement is based on risk of disease outbreaks and the consequences. For example, in the TB advantage programme the merit of bulls is released according to the performance of their progeny in bTB affected areas (Https://dairy.ahdb.org.uk/technical-information/breeding-genetics/tb-advantage/, 2016; Banos *et al.,* 2017). This information is more likely to be used by farmers in areas prone to experience bTB outbreaks.

Selecting for individual polymorphisms (such as in MAS) may have unintended consequences, whether for commercial, health or welfare traits. Generally, a variant intended for selection is tested in standard field conditions to ensure the candidate variant does not have unintended effects on economically significant traits (e.g. Dunkelberger et al., (2017)). However, under variable environmental situations such as in terms of response to other pathogens (beyond those tested) unintended consequences still present a risk. This risk may be realised in a range of ways such as on health and welfare (Twine, 2010), in terms of response to other infections (Khatib & Bormann, 2015) or in terms of reductions in genetic diversity (Dekkers & Hospital, 2002). Evaluation of the effect of the variant under the broad range of environmental conditions and the pathogens animals may be exposed to on farm, is not always feasible. This risk may partially be avoided by using a broader consideration of

robustness. As discussed in this thesis it may be possible to combine periods of elevated levels of reproductive failure (using the threshold method) to identify genetic signals which confer improved survivability regardless of the underlying aetiological basis of reproductive failure.

Whilst existing studies suggest that the well characterised WUR10000125 genotype (conferring improved resilience to one species of PRRSV) confers protection across different PRRSV strains (Hess et al., 2016); this may not hold for other loci or different varieties of PRRSV strains. This makes extrapolation of the effect of any favourable variant across pathogen strains / species impossible.

7.04 Future Considerations

Demands on food, and specifically food from livestock sources, are expected to continue to increase over the coming years driven by increased population, increased urbanisation and increases in meat consumption *per capita* as a result of increases in wealth in developing countries (McGilloway, 2005; Lutz & KC, 2010). These pressures may be further compounded by constraints placed on animal production industries to limit greenhouse gas emissions (Thornton, 2010). As livestock production intensifies to cope with these increased demands so to may disease prevalence, driven by increases in animal density and resulting impact on animal welfare (Hyun *et al.,* 1998; Maes *et al.,* 2000). Given the complex nature of animal response to external stressors and the possibility of compounded effects (Hyun *et al.,* 1998), the resilience and robustness of the animal across sources of stress is also likely to become more important in developing animals better adapted to more intensive production conditions.

7.05 Conclusion

The results from this study strongly imply that reproductive outcomes in response to PRRS can be improved by genetic control strategies. Specifically, that genetic variation in the

SSC4 region, previously identified as showing an effect on viral load and weight gain under PRRSV challenge, also has an effect on reproductive performance. According to this study other regions may also contain considerable effects on reproductive performance which may be employed in genomic or marker assisted selection strategies.

Bibliography

Adamo, S.A. (2004). How should behavioural ecologists interpret measurements of immunity? *Animal Behaviour*. 68 (6). p.pp. 1443–1449.

Adams, M.J., Elliot Lefkowitz, B.J., Andrew Q King, B.M., Harrison, R.L., Nick Knowles, B.J., Andrew Kropinski, B.M., Mart Krupovic, B., Kuhn, J.H. & Murilo Zerbini, F. (2016). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). *Archives of Virology*. 161 (10). p.pp. 2921–2949.

Adams, M.J., Lefkowitz, E.J., King, A.M.Q., Harrach, B., Harrison, R.L., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Mushegian, A.R., Nibert, M., Sabanadzovic, S., Sanfaçon, H., Siddell, S.G., Simmonds, P., Varsani, A., Zerbini, F.M., Gorbalenya, A.E. & Davison, A.J. (2017). Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Archives of Virology*. 162 (8). p.pp. 2505–2538.

Ait-Ali, T., Wilson, A.D., Westcott, D.G., Clapperton, M., Waterfall, M., Mellencamp, M.A., Drew, T.W., Bishop, S.C. & Archibald, A.L. (2007). Innate immune responses to replication of porcine reproductive and respiratory syndrome virus in isolated Swine alveolar macrophages. *Viral immunology*. 20 (1). p.pp. 105–18.

Akaike, H. (1974). *A New Look at the Statistical Model Identification*. In: Springer New York, pp. 215–222.

Albers, G.A.A., Gray, G.D., Piper, L.R., Barker, J.S.F., Jambre, L.F.L. & Barger, I.A. (1987). The genetics of resistance and resilience to Haemonchus contortus infection in young merino sheep. *International Journal for Parasitology*. 17 (7). p.pp. 1355–1363.

Albina, E., Carrat, C. & Charley, B. (1998). Interferon-alpha response to swine arterivirus (PoAV), the porcine reproductive and respiratory syndrome virus. *Journal of Interferon & Cytokine Research*. 18 (7). p.pp. 485–490.

Albina, E., Madec, F., Cariolet, R. & Torrison, J. (1994). Immune response and persistence of the porcine reproductive and respiratory syndrome virus in infected pigs and farm units. *The Veterinary record*. 134 (22). p.pp. 567–73.

Allende, R., Lewis, T.L., Lu, Z., Rock, D.L., Kutish, G.F., Ali, A., Doster, A.R. & Osorio, F.A. (1999). North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions. *Journal of General Virology*. 80 (2). p.pp. 307–315.

Alonso, C., Raynor, P.C., Davies, P.R., Torremorell, M., Núñez, A. & Clifford, D. (2015). Concentration, Size Distribution, and Infectivity of Airborne Particles Carrying Swine Viruses R. Tripp (ed.). *PLOS ONE*. 10 (8). p.p. e0135675.

Amadori, M. & Razzuoli, E. (2014). Immune Control of PRRS: Lessons to be Learned and Possible Ways Forward. *Frontiers in Veterinary Science*. 1. p.p. 2.

Amin, N., van Duijn, C.M., Aulchenko, Y.S.Y., Verny, C., Lemainque, A., Clerget-Darpoux, F., Thompson, E., Health, A., Martin, N., Montgomery, G., Goddard, M., Visscher, P., Majumder, P., Ghosh, S., Sing, C., Davignon, J., Isaacs, A., Sayed-Tabatabaei, F., Aulchenko, Y.S.Y., Zillikens, M., Sijbrands, E., Risch, N., Merikangas, K., Abecasis, G., Cardon, L., Cookson, W., Horvath, S., Xu, X., Lake, S., Silverman, E., Weiss, S., Lange, C., DeMeo, D., Laird, N., Boerwinkle, E., Chakraborty, R., Sing, C., Lange, K., Sinsheimer, J., Sobel, E., Havill, L., Dyer, T., Richardson, D., Mahaney, M., Blangero, J., Aulchenko, Y.S.Y., Koning, D., Haley, C., Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Devlin, B., Roeder, K., Bacanu, S., Devlin, B., Roeder, K., Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Clayton, D., Walker, N., Smyth, D., Pask, R., Cooper, J., Leutenegger, A., Prum, B., Genin, E., Verny, C., Lemainque, A., Gilmour, A., Gogel, B., Cullis, B., Welham, S., Thompson, R., Aulchenko, Y.S.Y., Ripke, S., Isaacs, A., Duijn, C. van, Pardo, L., MacKay, I., Oostra, B., Duijn, C. van, Aulchenko, Y.S.Y., Zollner, S., Pritchard, J., Durrant, C., Zondervan, K., Cardon, L., Hunt, S. & Deloukas, P. (2007). A Genomic Background Based Method for Association Analysis in Related Individuals P. Heutink (ed.). *PLoS ONE*. 2 (12). p.p. e1274.

Archibald, A., Audonnet, J.C., Babiuk, L., Bishop, S.C., Gay, C.G., McKay, J., Mallard, B., Plastow, G., Pinard van der Laan, M.H. & Torremorell, M. (2008). Animal genomics for animal health report: critical needs, problems to be solved, potential solutions, and a roadmap for moving forward. *Developments in biologicals*. 132. p.pp. 407–24.

Aulchenko, Y.S. (2011). Effects of Population Structure in Genome-wide Association Studies. In: *Analysis of Complex Disease Association Studies*. Amsterdam, Netherlands: Elsevier, pp. 123–156.

Aulchenko, Y.S., de Koning, D.-J. & Haley, C. (2007a). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*. 177 (1). p.pp. 577–85.

Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. (2007b). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23 (10). p.pp. 1294–1296.

Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press.

De Baere, M.I., Van Gorp, H., Delputte, P.L. & Nauwynck, H.J. (2012). Interaction of the European genotype porcine reproductive and respiratory syndrome virus (PRRSV) with sialoadhesin (CD169/Siglec-1) inhibits alveolar macrophage phagocytosis. *Veterinary Research*. 43 (1). p.p. 47.

Banos, G., Winters, M., Mrode, R., Mitchell, A.P., Bishop, S.C., Woolliams, J.A., Coffey, M.P., Pong-Wong, R., Matika, O., McDowell, S.W.J., Glass, E.J., Bishop, S.C., Rhodes, S., Rolfe, S., Sharp, M., Upton, H.M., Vordermeier, H.M., Watson, E., Welsh, W. & Whelan, A.O. (2017). Genetic evaluation for bovine tuberculosis resistance in dairy cattle. *Journal of dairy science*. 100 (2). p.pp. 1272–1281.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2016). Linear Mixed-Effects Models using 'Eigen' and S4. *Journal of Statistical Software*. 67 (1). p.pp. 1–48.

Baylis, M., Chihota, C., Stevenson, E., Goldmann, W., Smith, A., Sivam, K., Tongue, S. & Gravenor, M.B. (2004). Risk of scrapie in British sheep of different prion protein genotype. *Journal of General Virology*. 85 (9). p.pp. 2735–2740.

Beavis, W.D. (1998). QTL analysis: Power, precision, and accuracy. In: Paterson A H (ed.). *QTL analyses: power, precision, and accuracy,*. New York: CRC Press, pp. 145–162.

Benfield, D.A., Nelson, E., Collins, J.E., Harris, L., Goyal, S.M., Robison, D., Christianson, W.T., Morrison, R.B., Gorcyca, D. & Chladek, D. (1992). Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). *Journal of Veterinary Diagnostic Investigation*. 4 (2). p.pp. 127–133.

Bérénos, C., Ellis, P.A., Pilkington, J.G., Lee, S.H., Gratten, J. & Pemberton, J.M. (2015). Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*. 24 (8). p.pp. 1810–30.

Bierk, M.D., Dee, S.A., Rossow, K.D., Otake, S., Collins, J.E. & Molitor, T.W. (2001). Transmission of porcine reproductive and respiratory syndrome virus from persistently infected sows to contact controls. *Canadian Journal of Veterinary Research*. 65 (4). p.pp. 261–6.

Bishop, S.C. (2012). A consideration of resistance and tolerance for ruminant nematode infections. *Frontiers in genetics*. 3. p.p. 168.

Bishop, S.C., Doeschl-Wilson, A.B. & Woolliams, J.A. (2012). Uses and Implications of Field Disease Data for Livestock Genomic and Genetics Studies. *Frontiers in Genetics*. 3. p.p. 114.

Bishop, S.C. & MacKenzie, K.M. (2003). Genetic management strategies for controlling infectious diseases in livestock populations. *Genetics Selection Evolution*. 35 (Suppl. 1). p.pp. S3–S17.

Bishop, S.C. & Woolliams, J.A. (2014). Genomics and disease resistance studies in livestock. *Livestock science*. 166. p.pp. 190–198.

Bishop, S.C. & Woolliams, J.A. (2010). On the Genetic Interpretation of Disease Data S. A. Aziz (ed.). *PLoS ONE*. 5 (1). p.p. e8940.

Bloemraad, M., de Kluijver, E.P., Petersen, A., Burkhardt, G.E. & Wensvoort, G. (1994). Porcine reproductive and respiratory syndrome: temperature and pH stability of Lelystad virus and its survival in tissue specimens from viraemic pigs. *Veterinary Microbiology*. 42 (4). p.pp. 361–71.

Boddicker, N., Waide, E.H., Rowland, R.R.R., Lunney, J.K., Garrick, D.J., Reecy, J.M. & Dekkers, J.C.M. (2012). Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. *Journal of Animal Science*. 90 (6). p.pp. 1733–1746.

Boddicker, N.J., Bjorkquist, A., Rowland, R.R., Lunney, J.K., Reecy, J.M. & Dekkers, J.C. (2014a). Genome-wide association and genomic prediction for host response to porcine reproductive and respiratory syndrome virus infection. *Genetics Selection Evolution*. 46 (1). p.p. 18.

Boddicker, N.J., Garrick, D.J., Rowland, R.R.R., Lunney, J.K., Reecy, J.M. & Dekkers, J.C.M. (2014b). Validation and further characterization of a major quantitative trait locus associated with host response to experimental infection with porcine reproductive and respiratory syndrome virus. *Animal Genetics*. 45 (1). p.pp. 48–58.

Boerwinkle, E., Chakraborty, R. & Sing, C.F. (1986). The Use of Measured Genotype Information in the Analysis of Quantitative Phenotypes in Man. I. Models and analytical methods. *Annals of Human Genetics*. 50 (Pt 2). p.pp. 181–94.

Brogden, K.A. & Guthmiller, J.M. (2002). Porcine Respiratory Disease Complex. In: K. A. Brogden & J. M. Guthmiller (eds.). *Polymicrobial diseases*. Washington D.C.: ASM Press, pp. 231–258.

Burkard, C., Lillico, S.G., Reid, E., Jackson, B., Mileham, A.J., Ait-Ali, T., Whitelaw, C.B.A. & Archibald, A.L. (2017). Precision engineering for PRRSV resistance in pigs: Macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function D. R. Perez (ed.). *PLOS Pathogens*. 13 (2). p.p. e1006206.

Burton, D.R. (2002). Antibodies, viruses and vaccines. *Nature Reviews Immunology*. 2 (9). p.pp. 706–713.

Bush, W.S., Moore, J.H., Li, J., McDonnell, S. & Rabe, K. (2012). Chapter 11: Genome-Wide Association Studies F. Lewitter & M. Kann (eds.). *PLoS Computational Biology*. 8 (12). p.p. e1002822.

Butler, J.E., Lager, K.M., Golde, W., Faaberg, K.S., Sinkora, M., Loving, C. & Zhang, Y.I. (2014). Porcine reproductive and respiratory syndrome (PRRS): an immune dysregulatory pandemic. *Immunologic Research*. 59 (1–3). p.pp. 81–108.

Campbell, C.D., Ogburn, E.L., Lunetta, K.L., Lyon, H.N., Freedman, M.L., Groop, L.C., Altshuler, D., Ardlie, K.G. & Hirschhorn, J.N. (2005). Demonstrating stratification in a European American population. *Nature Genetics*. 37 (8). p.pp. 868–872.

Chen, N., Trible, B.R., Kerrigan, M.A., Tian, K. & Rowland, R.R.R. (2016). ORF5 of porcine reproductive and respiratory syndrome virus (PRRSV) is a target of diversifying selection as infection progresses from acute infection to virus rebound. *Infection, Genetics and Evolution*. 40. p.pp. 167–175.

Chen, W.-M. & Abecasis, G.R. (2007). Family-based association tests for genomewide association scans. *American journal of human genetics*. 81 (5). p.pp. 913–26.

Christianson, W.T. (1992). Stillbirths, Mummies, Abortions, and Early Embryonic Death. *Veterinary Clinics of North America: Food Animal Practice*. 8 (3). p.pp. 623–639.

Christopher-Hennings, J., Holler, L.D., Benfield, D.A. & Nelson, E.A. (2001). Detection and duration of porcine reproductive and respiratory syndrome virus in semen, serum, peripheral blood mononuclear cells, and tissues from Yorkshire, Hampshire, and Landrace boars. *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*. 13 (2). p.pp. 133–42.

Churchill, G.A. & Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*. 138 (3). p.pp. 963–71.

Collins, J.E., Benfield, D.A., Christianson, W.T., Harris, L., Hennings, J.C., Shaw, D.P., Goyal, S.M., McCullough, S., Morrison, R.B., Joo, H.S., Gorcyca, D. & Chladek, D. (1992). Isolation of Swine Infertility and Respiratory Syndrome Virus (Isolate ATCC VR-2332) in North America and Experimental Reproduction of the Disease in Gnotobiotic Pigs. *Journal of Veterinary Diagnostic Investigation*. 4 (2). p.pp. 117–126.

Conzelmann, K.-K., Visser, N., Van Woensel, P. & Thiel, H.-J. (1993). Molecular Characterization of Porcine Reproductive and Respiratory Syndrome Virus, a Member of the Arterivirus Group. *Virology*. 193 (1). p.pp. 329–339.

Darwich, L., Díaz, I. & Mateu, E. (2010). Certainties, doubts and hypotheses in porcine reproductive and respiratory syndrome virus immunobiology. *Virus Research*. 154 (1–2). p.pp. 123–132.

Davies, G., Genini, S., Bishop, S.C. & Giuffra, E. (2009). An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. *animal*. 3 (3). p.pp. 415–436.

Dawson, M., Moore, R.C. & Bishop, S.C. (2008). Progress and limits of PrP gene selection policy. *Veterinary Research*. 39 (4). p.p. 25.

Dee, S.A., Morrison, R.B. & }oo, H. (1993). Irradiating porcine reproductive and respiratory syndrome virus using two-site production and nursery depopulation. *Swine Health and Production*. 1 (5). p.pp. 20–23.

Dee, S., Deen, J., Rossow, K., Weise, C., Eliason, R., Otake, S., Joo, H.S. & Pijoan, C. (2003). Mechanical transmission of porcine reproductive and respiratory syndrome virus throughout a coordinated sequence of events during warm weather. *Canadian Journal of Veterinary Research*. 67 (1). p.pp. 12–9.

Dee, S., Deen, J., Rossow, K., Wiese, C., Otake, S., Joo, H.S. & Pijoan, C. (2002). Mechanical transmission of porcine reproductive and respiratory syndrome virus throughout a coordinated sequence of events during cold weather. *Canadian Journal of Veterinary Research*. 66 (4). p.pp. 232–9.

Dee, S., Otake, S., Oliveira, S. & Deen, J. (2009). Evidence of long distance airborne transport of porcine reproductive and respiratory syndrome virus and Mycoplasma hyopneumoniae. *Veterinary research*. 40 (4). p.p. 39.

Dekkers, J., Rowland, R.R.R., Lunney, J.K. & Plastow, G. (2017). Host genetics of response to porcine reproductive and respiratory syndrome in nursery pigs. *Veterinary Microbiology*. In Press (Corrected Proof).

Dekkers, J.C.M. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science*. 82 (13_suppl). p.pp. E313–E328.

Dekkers, J.C.M. & Hospital, F. (2002). Multifactorial Genetics: The use of Molecular Genetics in the Improvement of Agricultural Populations. *Nature Reviews Genetics*. 3 (1). p.pp. 22–32.

Devlin, B. & Roeder, K. (1999). Genomic Control for Association Studies. *Biometrics*. 55 (4). p.pp. 997–1004.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*. 418 (6898). p.pp. 700–707.

Doeschl-Wilson, A.B., Kyriazakis, I., Vincent, A., Rothschild, M.F., Thacker, E. & Galina-Pantoja, L. (2009). Clinical and pathological responses of pigs from two genetically diverse commercial lines to porcine reproductive and respiratory syndrome virus infection. *Journal of Animal Science*. 87 (5). p.pp. 1638–1647.

Doeschl-Wilson, A.B., Villanueva, B. & Kyriazakis, I. (2012). The first step toward genetic selection for host tolerance to infectious pathogens: obtaining the tolerance phenotype through group estimates. *Frontiers in genetics*. 3. p.p. 265.

Duan, X., Nauwynck, H.J., Favoreel, H.W. & Pensaert, M.B. (1998). Identification of a putative receptor for porcine reproductive and respiratory syndrome virus on porcine alveolar macrophages. *Journal of virology*. 72 (5). p.pp. 4520–3.

Duan, X., Nauwynck, H.J. & Pensaert, M.B. (1997a). Effects of origin and state of differentiation and activation of monocytes/macrophages on their susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV). *Archives of virology*. 142 (12). p.pp. 2483–97.

Duan, X., Nauwynck, H.J. & Pensaert, M.B. (1997b). Virus quantification and identification of cellular targets in the lungs and lymphoid tissues of pigs at different time intervals after inoculation with porcine reproductive and respiratory syndrome virus (PRRSV). *Veterinary Microbiology*. 56 (1–2). p.pp. 9–19.

Dunkelberger, J.R., Serão, N.V.L., Niederwerder, M.C., Kerrigan, M.A., Lunney, J.K., Rowland, R.R.R. & Dekkers, J.C.M. (2017). Effect of a major quantitative trait locus for porcine reproductive and respiratory syndrome (PRRS) resistance on response to coinfection with PRRS virus and porcine circovirus type 2b (PCV2b) in commercial pigs, with or without prior vaccination for PRRS1. *Journal of Animal Science*. 95 (2). p.pp. 584–598.

Dunn, O.J. (1959). Estimation of the Medians for Dependent Variables on JSTOR. *The Annals of Mathematical Statistics*. 30 (1). p.pp. 192–197.

Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*. 56 (293). p.p. 52.

Dwivedi, V., Manickam, C., Patterson, R., Dodson, K., Murtaugh, M., Torrelles, J.B., Schlesinger, L.S. & Renukaradhya, G.J. (2011). Cross-protective immunity to porcine reproductive and respiratory syndrome virus by intranasal delivery of a live virus vaccine with a potent adjuvant. *Vaccine*. 29 (23). p.pp. 4058–4066.

Dziuk, P.J. (1968). Effect of Number of Embryos and Uterine Space on Embryo Survival in the Pig. *Journal of Animal Science*. 27 (3). p.p. 673.

Ellis, S.A. & Hammond, J.A. (2014). The Functional Significance of Cattle Major Histocompatibility Complex Class I Genetic Diversity. *Annual Review of Animal Biosciences*. 2 (1). p.pp. 285–306.

Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to quantitative genetics*. 4th Ed. Burnt Mill, England: Longman.

Feng, S., Wang, S., Chen, C.-C. & Lan, L. (2011). GWAPower: a statistical power calculation software for genome-wide association studies with quantitative traits. *BMC Genetics*. 12 (1). p.p. 12.

Fisher, R.A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*. 52. p.pp. 399–433.

Forni, S., Aguilar, I. & Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*. 43 (1). p.p. 1.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N., Pato, M.T., Petryshen, T.L., Kolonel, L.N., Lander, E.S., Sklar, P., Henderson, B., Hirschhorn, J.N. & Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*. 36 (4). p.pp. 388–393.

Galina-Pantoja, L., Siggens, K., van Schriek, M.G.M. & Heuven, H.C.M. (2009). Mapping markers linked to porcine salmonellosis susceptibility. *Animal Genetics*. 40 (6). p.pp. 795–803.

Gangoso, L., Alcaide, M., Grande, J.M., Muñoz, J., Talbot, S.L., Sonsthagen, S.A., Sage, G.K. & Figuerola, J. (2012). Colonizing the world in spite of reduced MHC variation. *Journal of Evolutionary Biology*. 25 (7). p.pp. 1438–1447.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D. & Province, M.A. (2010). Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic Epidemiology*. 34 (1). p.pp. 100–5.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009). *ASREML 3.0*. Hemel Hempstead, UK, HP1 1ES: VSN International Ltd;

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2008). *ASReml User Guide Release 3.0*. Hemel Hempstead, HP1 1ES, UK: VSN International Ltd.

Goddard, M.E. & Hayes, B.J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*. 124 (6). p.pp. 323–330.

Gong, Y.-F., Lu, X., Wang, Z.-P., Hu, F., Luo, Y.-R., Cai, S.-Q., Qi, C.-M., Li, S., Niu, X.-Y., Qiu, X.-T., Zeng, J. & Zhang, Q. (2010). Detection of quantitative trait loci affecting haematological traits in swine via genome scanning. *BMC Genetics*. 11 (1). p.p. 56.

Van Gorp, H., Delputte, P.L. & Nauwynck, H.J. (2010). Scavenger receptor CD163, a Jack-of-all-trades and potential target for cell-directed therapy. *Molecular Immunology*. 47 (7–8). p.pp. 1650–1660.

Gradil, C., Dubuc, C. & Eaglesome, M.D. (1996). Porcine reproductive and respiratory syndrome virus: seminal transmission. *The Veterinary Record*. 138 (21). p.pp. 521–2.

Gredler, B., Fuerst, C. & Sölkner, J. (2007). Analysis of new fertility traits for the joint genetic evaluation in Austria and Germany. *Interbull Bulletin*. 37. p.pp. 152–155.

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C. & Schook, L.B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 491 (7424). p.pp. 393–398.

Hagenaars, T.J., Melchior, M.B., Bossers, A., Davidse, A., Engel, B. & van Zijderveld, F.G. (2010). Scrapie prevalence in sheep of susceptible genotype is declining in a population subject to breeding for resistance. *BMC Veterinary Research*. 6 (1). p.p. 25.

Halbur, P.G., Paul, P.S., Andrews, J.J., Sanderson, T.P., Ross, R.F., Schwartz, K.J., Frey, M.L., Erickson, B.J., Hill, H.T. & Hoffman, L.J. (1993). Experimental transmission of an apparent viral pneumonia in conventional and gnotobiotic pigs. *The Veterinary Record*. 132 (11). p.pp. 263–6.

Halbur, P.G., Paul, P.S., Frey, M.L., Landgraf, J., Eernisse, K., Meng, X.-J., Andrews, J.J., Lum, M.A. & Rathje, J.A. (1996). Comparison of the Antigen Distribution of Two US Porcine Reproductive and Respiratory Syndrome Virus Isolates with that of the Lelystad Virus. *Veterinary Pathology*. 33 (2). p.pp. 159–170.

Halbur, P.G., Rothschild, M.F., Thacker, B.J., Meng, X.-J., Paul, P.S. & Bruna, J.D. (1998). Differences in susceptibility of Duroc, Hampshire, and Meishan pigs to infection with a high-virulence strain (VR2385) of porcine reproductive and respiratory syndrome virus (PRRSV). *J. Anim. Breed. Genet*. 115 (1–6). p.pp. 181–189.

Han, K., Seo, H.W., Park, C., Oh, Y., Kang, I. & Chae, C. (2013). Comparative pathogenesis of type 1 (European genotype) and type 2 (North American genotype) porcine reproductive and respiratory syndrome virus in infected boar. *Virology journal*. 10. p.p. 156.

Harris, D.L. (Delbert L. (2000). *Multi-site pig production*. 1st Ed. Ames, Iowa 50014: Iowa State University Press.

Hartley, S., Gillund, F., van Hove, L., Wickson, F., Gorman, M. & Seager, T. (2016). Essential Features of Responsible Governance of Agricultural Biotechnology C. Marris (ed.). *PLOS Biology*. 14 (5). p.p. e1002453.

Hayes, B. & Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*. 53 (11). p.pp. 876–883.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. & Goddard, M.E. (2009a). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*. 41 (1). p.p. 51.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009b). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 92 (2). p.pp. 433–443.

Hayes, B.J., Visscher, P.M., Goddard, M.E., Verbyla, K., Goddard, M.E., Kvasz, A., Mni, M., Simon, P., Frère, J., Coppieters, W. & Georges, M. (2009c). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*. 91 (1). p.p. 47.

Henderson, C.. (1949). Estimation of changes in herd environment. *Journal of Dairy Science*. 32. p.p. 706.

Henderson, C.R. (1977). Best Linear Unbiased Prediction of Breeding Values Not in the Model for Records. *Journal of Dairy Science*. 60 (5). p.pp. 783–787.

Henderson, C.R. (1950). Estimation of genetic parameters. *The Annals of Mathematical Statistics*. 21 (2). p.pp. 309–310.

Henderson, C.R. (1973). Sire Evaluation and Genetic Trends. *Journal of Animal Science*. 1973 (Symposium). p.pp. 10–41.

Hess, A.S., Islam, Z., Hess, M.K., Rowland, R.R.R., Lunney, J.K., Doeschl-Wilson, A., Plastow, G.S. & Dekkers, J.C.M. (2016). Comparison of host genetic factors influencing pig response to infection with two North American isolates of porcine reproductive and respiratory syndrome virus. *Genetics Selection Evolution*. 48 (1). p.p. 43.

Hill, W.G. & Weir, B.S. (2011). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*. 93 (1). p.pp. 47–64.

Holm, B., Bakken, M., Klemetsdal, G. & Vangen, O. (2004). Genetic correlations between reproduction and production traits in swine. *Journal of Animal Science*. 82 (12). p.pp. 3458–64.

Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L. & Haley, C. a. (2013). Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *Journal of Swine Health and Production*. 21 (2). p.pp. 72–84.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J. & Bishop, S.C. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar). *Genetics*. 178 (2). p.pp. 1109–15.

Hoyos-Flight, M., Brady, E., Sang, H. & Whitelaw, B. (2017). Genome Editing and the Future of Farming meeting report. *Transgenic Research*. p.pp. 1–3.

Http://www.ed.ac.uk/roslin/research/isp/control-infectious-diseases/genetic-basis-of-host-resistance/breeding-tuberculosis-resistant-cattle (2017). *Breeding for tuberculosis resistant cattle*. 2017. The University of Edinburgh.

Https://dairy.ahdb.org.uk/technical-information/breeding-genetics/tb-advantage/ (2016). *TB Advantage - the genetics of bTB*: https://dairy.ahdb.org.uk/technical-information/breeding-genetics/tb-advantage/#.WXybmIjyvIU. 2016. Agriculture and Horticulture Development Board.

Hu, Z.-L., Dracheva, S., Jang, W., Maglott, D., Bastiaansen, J., Rothschild, M.F. & Reecy, J.M. (2005). A QTL resource and comparison tool for pigs: PigQTLDB. *Mammalian Genome*. 16 (10). p.pp. 792–800.

Hu, Z.-L., Park, C.A., Reecy, J.M., Nagarajan, R., Rice, T., Rao, D., Dwinell, M., Shimoyama, M., Eppig, J. & Reecy, J. (2016). Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research*. 44 (D1). p.pp. D827–D833.

Hughes, A.L. & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America*. 86 (3). p.pp. 958–62.

Hutchinson, A.B., Farnham, P.G., Dean, H.D., Ekwueme, D.U., del Rio, C., Kamimoto, L. & Kellerman, S.E. (2006). The Economic Burden of HIV in the United States in the Era of Highly Active Antiretroviral Therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 43 (4). p.pp. 451–457.

Hyun, Y., Ellis, M., Riskowski, G. & Johnson, R.W. (1998). Growth performance of pigs subjected to multiple concurrent environmental stressors. *Journal of Animal Science*. 76 (3). p.p. 721.

Idexx Herdchek (2003). *IDEXX Production Animal Services Exposure Responses Porcine Reproductive and Respiratory Syndrome Virus*. Main.

Iles, M.M. & Barrett, J.H. (2011). Single-locus Tests of Association for Population-based Studies. In: *Analysis of Complex Disease Association Studies*. Elsevier, pp. 109–122.

Illumina (2015). *PorcineSNP60 v2 Genotyping BeadChip*: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_porcinesnp60.pdf. 2015.

Islam, Z.U., Bishop, S.C., Savill, N.J., Rowland, R.R.R., Lunney, J.K., Trible, B. & Doeschl-Wilson, A.B. (2013). Quantitative analysis of porcine reproductive and respiratory syndrome (PRRS) viremia profiles from experimental infection: A statistical modelling approach. *PLoS ONE*. 8 (12). p.pp. 1–13.

Johnson, R.K., Nielsen, M.K. & Casey, D.S. (1999). Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. *Journal of animal science*. 77 (3). p.pp. 541–57.

Johnston, L.J., Renteria, A. & Hannon, M.R. (2003). Improving validity of on-farm research. *Journal of Swine Health and Production — J Swine Health Prod*. 11 (5). p.pp. 240–246.

Karlsson, L.J.E., Greeff, J.C., Eady, S., Greeff, J., Karlsson, J., Morcombe, P., Dalton–Morgan, G. & Roberts, D. (2006). Selection response in fecal worm egg counts in the Rylington Merino parasite resistant flock. *Australian Journal of Experimental Agriculture*. 46 (7). p.p. 809.

Karniychuk, U.U. & Nauwynck, H.J. (2013). Pathogenesis and prevention of placental and transplacental porcine reproductive and respiratory syndrome virus infection. *Veterinary research*. 44 (1). p.p. 95.

Karniychuk, U.U., Saha, D., Geldhof, M., Vanhee, M., Cornillie, P., Van den Broeck, W. & Nauwynck, H.J. (2011). Porcine reproductive and respiratory syndrome virus (PRRSV) causes apoptosis during its replication in fetal implantation sites. *Microbial Pathogenesis*. 51 (3). p.pp. 194–202.

Karniychuk, U.U., Saha, D., Vanhee, M., Geldhof, M., Cornillie, P., Caij, A.B., De Regge, N. & Nauwynck, H.J. (2012). Impact of a novel inactivated PRRS virus vaccine on virus replication and virus-induced pathology in fetal implantation sites and fetuses upon challenge. *Theriogenology*. 78 (7). p.pp. 1527–1537.

Keffaber, K.K. (1989). Reproductive failure of unknown etiology. *American Association of Swine Practitioners Newsletter*. 1. p.pp. 1–10.

Kennedy, B.W., Quinton, M. & van Arendonk, J.A. (1992). Estimation of effects of single genes on quantitative traits. *Journal of Animal Science*. 70 (7). p.pp. 2000–12.

Khatib, H. & Bormann, J.M. (2015). *Molecular and Quantitative Animal Genetics*. First. Wiley-Blackwell: Hoboken, New Jersey.

Kleinbaum, D.G., Kupper, L.L., Nizam, A. & Rosenberg, E.S. (2008). *Applied regression analysis and other multivariable methods*. 5th Ed. Boston : Cengage Learning.

Knap, P.W. (2005). Breeding robust pigs. *Australian Journal of Experimental Agriculture*. 45 (8). p.p. 763.

Knol, E.., Leenhouwers, J.. & van der Lende, T. (2002). Genetic aspects of piglet survival. *Livestock Production Science*. 78 (1). p.pp. 47–55.

Koltes, J.E., Fritz-Waters, E., Eisley, C.J., Choi, I., Bao, H., Kommadath, A., Serão, N.V.L., Boddicker, N.J., Abrams, S.M., Schroyen, M., Loyd, H., Tuggle, C.K., Plastow, G.S., Guan, L., Stothard, P., Lunney, J.K., Liu, P., Carpenter, S., Rowland, R.R.R., Dekkers, J.C.M. & Reecy, J.M. (2015). Identification of a putative quantitative trait nucleotide in guanylate binding protein 5 for host response to PRRS virus infection. *BMC genomics*. 16 (1). p.p. 412.

Komijn, R.E., van der Sande, W.J.H. & van Klink, E.G.M. (1991). Report on the epidemiology of PRRS in The Netherlands. In: *Proceedings Seminar on Porcine Reproductive and Respiratory Syndrome*. 1991, Brussels, Belgium., pp. 8–12.

Krapp, C., Hotter, D., Gawanbacht, A., McLaren, P., Kluge, S., St?rzel, C., Mack, K., Reith, E., Engelhart, S., Ciuffi, A., Hornung, V., Sauter, D., Telenti, A. & Kirchhoff, F. (2016). Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. *Cell Host & Microbe*. 19 (4). p.pp. 504–514.

Kreft, I. & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London EC1Y 1SP: SAGE Publications, Ltd.

Kristensen, C.S., Bøtner, A., Angen, Ø., Sørensen, V., Jorsal, S.E.L., Takai, H., Barfod, K. & Nielsen, J.P. (2002). Airborne transmission of A. pleuropneumoniae and PRRS virus between pig units. In: *Proceedings of the International Pig Veterinary Society Congress*. 2002, Iowa: Iowa State University Press, p. 272.

Kuhn, J.H., Lauck, M., Bailey, A.L., Shchetinin, A.M., Vishnevskaya, T. V., Bào, Y., Ng, T.F.F., LeBreton, M., Schneider, B.S., Gillis, A., Tamoufe, U., Diffo, J.L.D., Takuo, J.M., Kondov, N.O., Coffey, L.L., Wolfe, N.D., Delwart, E., Clawson, A.N., Postnikova, E., Bollinger, L., Lackemeyer, M.G., Radoshitzky, S.R., Palacios, G., Wada, J., Shevtsova, Z. V., Jahrling, P.B., Lapin, B.A., Deriabin, P.G., Dunowska, M., Alkhovsky, S. V., Rogers, J., Friedrich, T.C., O'Connor, D.H. & Goldberg, T.L. (2016). Reorganization and expansion of the nidoviral family Arteriviridae. *Archives of Virology*. 161 (3). p.pp. 755–768.

Kuznetsova, A., Brockhoff, B. & Christensen, H.B. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. Vienna, Austria: R Foundation for Statistical Computing.

Ladinig, A., Ashley, C., Detmer, S.E., Wilkinson, J.M., Lunney, J.K., Plastow, G. & Harding, J.C.S. (2015a). Maternal and fetal predictors of fetal viral load and death in third trimester, type 2 porcine reproductive and respiratory syndrome virus infected pregnant gilts. *Veterinary research*. 46. p.p. 107.

Ladinig, A., Detmer, S.E., Clarke, K., Ashley, C., Rowland, R.R.R., Lunney, J.K. & Harding, J.C.S. (2015b). Pathogenicity of three type 2 porcine reproductive and respiratory syndrome virus strains in experimentally inoculated pregnant gilts. *Virus Research*. 203. p.pp. 24–35.

Ladinig, A., Foxcroft, G., Ashley, C., Lunney, J.K., Plastow, G. & Harding, J.C.S. (2014a). Birth Weight, Intrauterine Growth Retardation and Fetal Susceptibility to Porcine Reproductive and Respiratory Syndrome Virus. *PLoS ONE*. 9 (10). p.p. e109541.

Ladinig, A., Gerner, W., Saalmüller, A., Lunney, J.K., Ashley, C. & Harding, J.C. (2014b). Changes in leukocyte subsets of pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. *Veterinary Research*. 45 (1). p.p. 128.

Ladinig, A., Lunney, J.K., Souza, C.J., Ashley, C., Plastow, G. & Harding, J.C. (2014c). Cytokine profiles in pregnant gilts experimentally infected with porcine reproductive and respiratory syndrome virus and relationships with viral load and fetal outcome. *Veterinary Research*. 45 (1). p.p. 113.

Ladinig, A., Wilkinson, J., Ashley, C., Detmer, S.E., Lunney, J.K., Plastow, G. & Harding, J.C.S. (2014d). Variation in fetal outcome, viral load and ORF5 sequence mutations in a large scale study of phenotypic responses to late gestation exposure to type 2 porcine reproductive and respiratory syndrome virus. *PLoS ONE*. 9 (4). p.pp. 1–11.

LeFebvre, R. (2015). Fetal mummification in the major domestic species: current perspectives on causes and management. *Veterinary Medicine: Research and Reports*. 6. p.p. 233.

Lenth, R. V. (2016). Least-Squares Means: The R Package (lsmeans). *Journal of Statistical Software*. 69 (1). p.pp. 1–33.

Lewis, C.R.G. (2008). *Dissecting the Genetic Control of Resistance to Porcine Reproductive and Respiratory Syndrome (PRRS)*. The University of Edinburgh.

Lewis, C.R.G., Ait-Ali, T., Clapperton, M., Archibald, A.L. & Bishop, S.C. (2007). Genetic perspectives on host responses to porcine reproductive and respiratory syndrome (PRRS). *Viral Immunol*. 20 (3). p.pp. 343–358.

Lewis, C.R.G., Torremorell, M. & Bishop, S.C. (2009a). Effects of Porcine Reproductive and Respiratory Syndrome (PRRS) virus infection on the performance of commercial sows and gilts of different parities and lines. *J Swine Health and Production*. 17 (3). p.pp. 140–147.

Lewis, C.R.G., Torremorell, M., Galina-Pantoja, L. & Bishop, S.C. (2009b). Genetic parameters for performance traits in commercial sows estimated before and after an outbreak of porcine reproductive and respiratory syndrome. *Journal of Animal Science*. 87 (3). p.pp. 876–884.

Lewis, C.R.G., Torremorell, M., Galina-Pantoja, L., Deeb, N., Mellencamp, M.A., Archibald, A.L. & Bishop, S.C. (2009c). A Genome-Wide Association Analysis Identifying SNPs for PRRS Tolerance on a Commercial Pig Farm. In: *Proceedings Association for the Advancement of Animal Breeding and Genetics*. 2009, Barossa Valley, South Australia, pp. 187–190.

Lewontin, R.C. & Kojima, K. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution*. 14 (4). p.p. 458.

Linhares, D.C.L., Johnson, C., Morrison, R.B., Neumann, E., Kliebenstein, J., Johnson, C., Mabry, J., Bush, E., Seitzinger, A., Holtkamp, D., Kliebenstein, J., Neumann, D., Nieuwenhuis, N., Duinhof, T., Nes, A. van, Corzo, C., Mondaca, E., Wayne, S., Torremorell, M., Dee, S., Davies, P., Torremorell, M., Henry, S., Christianson, W., Desrosiers, R., Boutin, M., Linhares, D., Cano, J., Torremorell, M., Morrison, R., Chaminuka, P., McCrindle, C., Udo, H., Li, M., Robinson, E., Oberle, D., Lucas, P., Rodgers, J., Bird, S., Larson, J., DiLorenzo, N., Dahlen, C., DiCostanzo, A., Sami, A., Abdelrahman, M., El-Asheeri, A., Radwan, M., Abbitt, B., Townsend, K., Metropolis, N., Ulam, S., Dewey, C., Wilson, S., Buck, P., Leyenaar, J., Nielsen, J., Bøtner, A., Bille-Hansen, V., Oleksiewicz, M., Storgaard, T., Velayudhan, D., Kim, I., Nyachoti, C., Murtaugh, M., Genzow, M., Arruda, A., Friendship, R., Carpenter, J., Hand, K., Ojkic, D., Poljak, Z., Bøtner, A., Strandbygaard, B., Sørensen, K., Have, P., Madsen, K., Madsen, E., Dewey, C., Wilson, S., Buck, P. & Leyenaar, J. (2015). Economic Analysis of Vaccination Strategies for PRRS Control F. C. Leung (ed.). *PLOS ONE*. 10 (12). p.p. e0144265.

Liu, L., Zhang, D., Liu, H. & Arendt, C. (2013). Robust methods for population stratification in genome wide association studies. *BMC Bioinformatics*. 14 (1). p.p. 132.

Lough, G., Rashidi, H., Kyriazakis, I., Dekkers, J.C.M., Hess, A., Hess, M., Deeb, N., Kause, A., Lunney, J.K., Rowland, R.R.R., Mulder, H.A. & Doeschl-Wilson, A. (2017). Use of multi-trait and random regression models to identify genetic variation in tolerance to porcine reproductive and respiratory syndrome virus. *Genetics Selection Evolution*. 49 (1). p.p. 37.

Loving, C.L., Osorio, F.A., Murtaugh, M.P. & Zuckermann, F.A. (2015). Innate and adaptive immunity against Porcine Reproductive and Respiratory Syndrome Virus. *Veterinary Immunology and Immunopathology*. 167 (1–2). p.pp. 1–14.

Lunney, J.K. & Chen, H. (2010). Genetic control of porcine reproductive and respiratory syndrome virus responses. *Virus Research*. 154 (1–2). p.pp. 1–6.

Lunney, J.K., Fang, Y., Ladinig, A., Chen, N., Li, Y., Rowland, B. & Renukaradhya, G.J. (2016). Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): Pathogenesis and Interaction with the Immune System. *Annual Review of Animal Biosciences*. 4. p.pp. 129–154.

Lunney, J.K., Steibel, J.P., Reecy, J.M., Fritz, E., Rothschild, M.F. & Kerrigan, M. (2011). Probing genetic control of swine responses to PRRSV infection: current progress of the PRRS host genetics consortium. *BMC Proc*. 5 ((Suppl. 4)). p.pp. 1–5.

Lush, J.L. (1937). *Animal Breeding Plans*. Second Edi. Aimes, Iowa: Iowa State College Press.

Lutz, W. & KC, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 365 (1554). p.pp. 2779–27791.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.

MacKenzie, K. (1999). *Quantifying Selection for Resistance to Infectious Diseases in Pigs using Genetic Epidemiological Models*. University of Edinburgh.

MacKenzie, K. & Bishop, S.C. (1999). A discrete-time epidemiological model to quantify selection for disease resistance. *Animal Science*. 69 (3). p.pp. 543–551.

Maes, D., Deluyker, H., Verdonck, M., Castryck, F., Miry, C., Vrijens, B. & de Kruif, A. (2000). Herd factors associated with the seroprevalences of four major respiratory pathogens in slaughter pigs from farrow-to-finish pig herds. *Veterinary Research*. 31 (3). p.pp. 313–327.

Makgahlela, M.L., Strandén, I., Nielsen, U.S., Sillanpää, M.J. & Mäntysaari, E.A. (2014). Using the unified relationship matrix adjusted by breed-wise allele frequencies in genomic evaluation of a multibreed population. *Journal of Dairy Science*. 97 (2). p.pp. 1117–1127.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A. & Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*. 461 (7265). p.pp. 747–753.

Mardassi, H., Mounir, S. & Dea, S. (1995). Molecular analysis of the ORFs 3 to 7 of porcine reproductive and respiratory syndrome virus, Quebec reference strain. *Archives of Virology*. 140 (8). p.pp. 1405–1418.

Mccaw, M.B. (2000). Effect of reducing crossfostering at birth on piglet mortality and performance during an acute outbreak of porcine reproductive and respiratory syndrome. *Swine Health and Production*. 8 (1). p.pp. 15–21.

McGilloway, D.A. (2005). *Grassland : a global resource*. Internatio. Wageningen: Wageningen Academic Publishers.

Mendel, G.J. (1865). Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn*. IV. p.pp. 3–47.

Meng, X.J. (2000). Heterogeneity of porcine reproductive and respiratory syndrome virus: implications for current vaccine efficacy and future vaccine development. *Veterinary Microbiology*. 74 (4). p.pp. 309–329.

Meng, X.J., Paul, P.S., Halbur, P.G. & Lum, M.A. (1995). Phylogenetic analyses of the putative M (ORF 6) and N (ORF 7) genes of porcine reproductive and respiratory syndrome virus (PRRSV): implication for the existence of two genotypes of PRRSV in the U.S.A. and Europe. *Archives of Virology*. 140 (4). p.pp. 745–55.

Mengeling, W.L., Lager, K.M. & Vorwald, A.C. (1994). Temporal characterization of transplacental infection of porcine fetuses with porcine reproductive and respiratory syndrome virus. *American Journal of Veterinary Research*. 55 (10). p.pp. 1391–8.

Meuwissen, T. & Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*. 24 (4). p.p. 305.

Meuwissen, T.H., Hayes, B.J. & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157 (4). p.pp. 1819–29.

Moen, T., Baranski, M., Sonesson, A.K. & Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC Genomics*. 10 (1). p.p. 368.

Moen, T., Torgersen, J., Santi, N., Davidson, W.S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., Lubieniecki, K.P., Isdal, E. & Lien, S. (2015). Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics*. 200 (4). p.pp. 1313–1326.

Morgan, T.H. (1911). Random Segregation Versus Coupling In Mendelian Inheritance. *Science*. 34 (873). p.pp. 384–384.

Mortensen, S. & Madsen, K.S. (1992). The occurrence of PRRS in Denmark. *American Association of Swine Practitioners Newsletter*. 4 (4). p.p. 48.

Mortensen, S., Stryhn, H., Søgaard, R., Boklund, A., Stärk, K.D.C., Christensen, J. & Willeberg, P. (2002). Risk factors for infection of sow herds with porcine reproductive and respiratory syndrome (PRRS) virus. *Preventive Veterinary Medicine*. 53 (1–2). p.pp. 83–101.

Murtaugh, M.P., Elam, M.R. & Kakach, L.T. (1995). Comparison of the structural protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus. *Archives of Virology*. 140 (8). p.pp. 1451–1460.

Murtaugh, M.P. & Genzow, M. (2011). Immunological solutions for treatment and prevention of porcine reproductive and respiratory syndrome (PRRS). *Vaccine*. 29 (46). p.pp. 8192–8204.

Murtaugh, M.P., Xiao, Z. & Zuckermann, F. (2002). Immunological Responses of Swine to Porcine Reproductive and Respiratory Syndrome Virus Infection. *Viral Immunology*. 15 (4). p.pp. 533–547.

Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A.A., Pramstaller, P.P., Hastie, N., Wright, A.F. & Haley, C.S. (2012). Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping S. Moore (ed.). *PLoS ONE*. 7 (10). p.p. e46501.

Nelsen, C.J., Murtaugh, M.P. & Faaberg, K.S. (1999). Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents. *Journal of Virology*. 73 (1). p.pp. 270–80.

Neyman, J. & Pearson, E.S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Mathematical or Physical Character*. 231. p.pp. 289–337.

Nodelijk, G. (2002). Porcine Reproductive and Respiratory Syndrome (PRRS) with special reference to clinical aspects and diagnosis: A review. *Veterinary Quarterly*. 24 (2). p.pp. 95–100.

Nodelijk, G., Nielen, M., De Jong, M.C.. & Verheijden, J.H.. (2003). A review of porcine reproductive and respiratory syndrome virus in Dutch breeding herds: population dynamics and clinical relevance. *Preventive Veterinary Medicine*. 60 (1). p.pp. 37–52.

Okamura, T., Onodera, W., Tayama, T., Kadowaki, H., Kojima-Shibata, C., Suzuki, E., Uemoto, Y., Mikawa, S., Hayashi, T., Awata, T., Fujishima-Kanaya, N., Mikawa, A., Uenishi, H. & Suzuki, K. (2012). A genome-wide scan for quantitative trait loci affecting respiratory disease and immune capacity in Landrace pigs. *Animal Genetics*. 43 (6). p.pp. 721–9.

Olanratmanee, E., Wongyanin, P., Thanawongnuwech, R. & Tummaruk, P. (2015). Prevalence of porcine reproductive and respiratory syndrome virus detection in aborted fetuses, mummified fetuses and stillborn piglets using quantitative polymerase chain reaction. *The Journal of Veterinary Medical Science*. 77 (9). p.pp. 1071–7.

Orrett, C.M., Deeb, N., Pong-Wong, R., Matika, O., Lewis, C.R.G., McLaren, D.G., Archibald, A. & Bishop, S. (2014). Regional Heritability Mapping of Production Traits in Epidemic Porcine Reproductive and Respiratory Syndrome. In: *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*. 2014, Vancouver, p. 100.

Orrett, C.M., Matika, O., Archibald, A., Lewis, C.R.G., Mclaren, D., Deeb, N. & Bishop, S. (2013). Genetic of host response to infection with porcine reproductive and respiratory syndrome virus (PRRSv). *Advances in Animal Bioscience*. 4. p.p. 81.

Otake, S., Dee, S.A., Jacobson, L., Torremorell, M. & Pijoan, C. (2002a). Evaluation of aerosol transmission of porcine reproductive and respiratory syndrome virus under controlled field conditions. *The Veterinary Record*. 150 (26). p.pp. 804–8.

Otake, S., Dee, S.A., Rossow, K.D., Joo, H.S., Deen, J., Molitor, T.W. & Pijoan, C. (2002b). Transmission of porcine reproductive and respiratory syndrome virus by needles. *The Veterinary Record*. 150 (4). p.pp. 114–5.

Otake, S., Dee, S.A., Rossow, K.D., Moon, R.D. & Pijoan, C. (2002c). Mechanical transmission of porcine reproductive and respiratory syndrome virus by mosquitoes, Aedes vexans (Meigen). *Canadian Journal of Veterinary Research*. 66 (3). p.pp. 191–5.

Otake, S., Dee, S.A., Rossow, K.D., Moon, R.D., Trincado, C. & Pijoan, C. (2003). Transmission of porcine reproductive and respiratory syndrome virus by houseflies (Musca domestica). *The Veterinary Record*. 152 (3). p.pp. 73–6.

Otake, S., Dee, S., Corzo, C., Oliveira, S. & Deen, J. (2010). Long-distance airborne transport of infectious PRRSV and Mycoplasma hyopneumoniae from a swine population infected with multiple viral variants. *Veterinary Microbiology*. 145 (3–4). p.pp. 198–208.

Patterson, H.D. & Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*. 58 p.pp. 545–554.

de Paz, X. (2015). *Xavier de Paz - pig333, pig to pork community*: https://www.pig333.com/articles/prrs-cost-for-the-european-swine-industry_10069/. 2015. pig333.

Petry, D.B., Holl, J.W., Weber, J.S., Doster, A.R., Osorio, F.A. & Johnson, R.K. (2005). Biological responses to porcine respiratory and reproductive syndrome virus in pigs of two genetic populations. *Journal of animal science*. 83 (7). p.pp. 1494–502.

Petry, D.B., Lunney, J., Boyd, P., Kuhar, D., Blankenship, E. & Johnson, R.K. (2007). Differential immunity in pigs with high and low responses to porcine reproductive and respiratory syndrome virus infection. *Journal of Animal Science*. 85 (9). p.p. 2075.

Pitkin, A., Deen, J. & Dee, S. (2009). Further assessment of fomites and personnel as vehicles for the mechanical transport and transmission of porcine reproductive and respiratory syndrome virus. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire*. 73 (4). p.pp. 298–302.

Pittman, J.S. (2008). Reproductive failure associated with porcine circovirus type 2 in gilts. *Journal of Swine Health and Production*. 16 (3). p.pp. 144–148.

Pol, J.M.A., van Dijk, J.E., Wensvoort, G. & Terpstra, C. (1991). Pathological, ultrastructural, and immunohistochemical changes caused by Lelystad virus in experimentally induced infections of mystery swine disease (synonym: Porcine epidemic abortion and respiratory syndrome (PEARS)). *Veterinary Quarterly*. 13 (3). p.pp. 137–143.

Powell, J.E., Visscher, P.M. & Goddard, M.E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*. 11 (11). p.pp. 800–805.

Prather, R.S., Rowland, R.R.R., Ewen, C., Trible, B., Kerrigan, M., Bawa, B., Teson, J.M., Mao, J., Lee, K., Samuel, M.S., Whitworth, K.M., Murphy, C.N., Egen, T. & Green, J.A. (2013). An intact sialoadhesin (Sn/SIGLEC1/CD169) is not required for attachment/internalization of the porcine reproductive and respiratory syndrome virus. *Journal of Virology*. 87 (17). p.pp. 9538–46.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 38 (8). p.pp. 904–909.

Prieto, C., Sánchez, R., Martín-Rillo, S., Suárez, P., Simarro, I., Solana, A. & Castro, J.M. (1996). Exposure of gilts in early gestation to porcine reproductive and respiratory syndrome virus. *The Veterinary Record*. 138 (22). p.pp. 536–9.

Prieto, C., Suárez, P., Simarro, I., García, C., Fernández, A. & Castro, J.M. (1997a). Transplacental infection following exposure of gilts to porcine reproductive and respiratory syndrome virus at the onset of gestation. *Veterinary Microbiology*. 57 (4). p.pp. 301–11.

Prieto, C., Suárez, P., Simarro, I., García, C., Martín-Rillo, S. & Castro, J.M. (1997b). Insemination of susceptible and preimmunized gilts with boar semen containing porcine reproductive and respiratory syndrome virus. *Theriogenology*. 47 (3). p.pp. 647–54.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahe, M.C. & Murtaugh, M.P. (2017). Mechanisms of Adaptive Immunity to Porcine Reproductive and Respiratory Syndrome Virus. *Viruses*. 9 (6).

Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., Hansen, M.S., Hedegaard, J., Hu, Z.-L., Kerstens, H.H., Law, A.S., Megens, H.-J., Milan, D., Nonneman, D.J., Rohrer, G.A., Rothschild, M.F., Smith, T.P.L., Schnabel, R.D., Van Tassell, C.P., Taylor, J.F., Wiedmann, R.T., Schook, L.B. & Groenen, M.A.M. (2009). Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology L. Orban (ed.). *PLoS ONE*. 4 (8). p.p. e6524.

Rashidi, H., Mulder, H.A., Mathur, P., van Arendonk, J.A.M. & Knol, E.F. (2014). Variation among sows in response to porcine reproductive and respiratory syndrome. *Journal of Animal Science*. 92 (1). p.pp. 95–105.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models : applications and data analysis methods*. Sage Publications.

Van Reeth, K., Labarque, G., Nauwynck, H. & Pensaert, M. (1999). Differential production of proinflammatory cytokines in the pig lung during different respiratory virus infections: correlations with pathogenicity. *Research in Veterinary Science*. 67 (1). p.pp. 47–52.

Van Regenmortel, M.H. (1989). Applying the species concept to plant viruses. *Archives of Virology*. 104 (1–2). p.pp. 1–17.

Reich, D.E. & Goldstein, D.B. (2001). Detecting association in a case-control study while correcting for population stratification. *Genetic Epidemiology*. 20 (1). p.pp. 4–16.

Reiner, G., Bertsch, N., Hoeltig, D., Selke, M., Willems, H., Gerlach, G.F., Tuemmler, B., Probst, I., Herwig, R., Drungowski, M. & Waldmann, K.H. (2014). Identification of QTL affecting resistance/susceptibility to acute Actinobacillus pleuropneumoniae infection in swine. *Mammalian genome : Official Journal of the International Mammalian Genome Society*. 25 (3–4). p.pp. 180–91.

Reiner, G., Willems, H., Pesch, S. & Ohlinger, V.F. (2010). Variation in resistance to the *Porcine Reproductive and Respiratory Syndrome Virus (PRRSV)* in Pietrain and Miniature pigs. *Journal of Animal Breeding and Genetics*. 127 (2). p.pp. 100–106.

Renukaradhya, G.J., Alekseev, K., Jung, K., Fang, Y. & Saif, L.J. (2010). Porcine Reproductive and Respiratory Syndrome Virus–Induced Immunosuppression Exacerbates the Inflammatory Response to Porcine Respiratory Coronavirus in Pigs. *Viral Immunology*. 23 (5). p.pp. 457–466.

Riggio, V., Pong-Wong, R., Sallé, G., Usai, M.G., Casu, S., Moreno, C.R., Matika, O. & Bishop, S.C. (2014). A joint analysis to identify loci underlying variation in nematode resistance in three European sheep populations. *Journal of Animal Breeding and Genetics*. 131 (6). p.pp. 426–436.

Rossow, K.D. (1998). Porcine reproductive and respiratory syndrome. *Veterinary Pathology*. 35 (1). p.pp. 1–20.

Rothschild, M.F. & Ruvinsky, A. (2011). *The genetics of the pig*. Second. Wallingford: CABI.

Rueff, L. (2000). Diagnostic approaches to reproductive failure in pigs. *Journal of Swine Health and Production*. 8 (6). p.pp. 285–287.

Schneider, J.F., Rempel, L.A., Snelling, W.M., Wiedmann, R.T., Nonneman, D.J. & Rohrer, G.A. (2012). Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data. *Journal of Animal Science*. 90 (10). p.pp. 3360–7.

Schook, L.B., Beever, J.E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., Milan, D., Rohrer, G. & Eversole, K. (2005). Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Comparative and functional genomics*. 6 (4). p.pp. 251–5.

Schroyen, M., Eisley, C., Koltes, J.E., Fritz-Waters, E., Choi, I., Plastow, G.S., Guan, L., Stothard, P., Bao, H., Kommadath, A., Reecy, J.M., Lunney, J.K., Rowland, R.R.R., Dekkers, J.C.M. & Tuggle, C.K. (2016). Bioinformatic analyses in early host response to Porcine Reproductive and Respiratory Syndrome virus (PRRSV) reveals pathway differences between pigs with alternate genotypes for a major host response QTL. *BMC Genomics*. 17 (1). p.p. 196.

Schukken, Y., De Jong, M. & Komijn, R. (1992). Diagnosis Porcine Epidemic Abortion and Respiratory Syndrome (PEARS)-positive and PEARS-negative for breeding and multiplier herds in the Netherlands using statistical methods. *Tijdschrift voor Diergeneeskunde*. 117. p.pp. 259–264.

Self, S.G. & Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*. 82 (398). p.pp. 605–610.

Serão, N., Matika, O., Kemp, R. a., Harding, J.C.S., Bishop, S.C., Plastow, G.S. & Dekkers, J.C.M. (2014). Genetic analysis of reproductive traits and antibody response in a PRRS outbreak herd. *Journal of Animal Science*. p.pp. 2905–2921.

Serão, N.V.L., Kemp, R.A., Mote, B.E., Willson, P., Harding, J.C.S., Bishop, S.C., Plastow, G.S. & Dekkers, J.C.M. (2016). Genetic and genomic basis of antibody response to porcine reproductive and respiratory syndrome (PRRS) in gilts and sows. *Genetics Selection Evolution*. 48 (1). p.p. 51.

Serenius, T., Sevón-aimonen, M. -l., Kause, A., Mäntysaari, E.A. & Mäki-tanila, A. (2004). Selection potential of different prolificacy traits in the finnish landrace and large white populations. *Acta Agriculturae Scandinavica, Section A - Animal Science*. 54 (1). p.pp. 36–43.

Sevón-Aimonen, M.-L. & Uimari, P. (2013). Heritability of sow longevity and lifetime prolificacy in Finnish Yorkshire and Landrace pigs. *Agricultural and Food Science*. 22 (3). p.pp. 325–330.

Shin, J.-H., Blay, S., Graham, J. & McNeney, B. (2006). LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *Journal of Statistical Software*. 16 (Code Snippet 3). p.pp. 1–9.

Shirali, M., Pong-Wong, R., Navarro, P., Knott, S., Hayward, C., Vitart, V., Rudan, I., Campbell, H., Hastie, N.D., Wright, A.F. & Haley, C.S. (2016). Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. *Heredity*. 116 (3). p.pp. 333–338.

Simms, E.L. & Triplett, J. (1994). Costs and Benefits of Plant Responses to Disease: Resistance and Tolerance. *Evolution*. 48 (6). p.p. 1973.

Soetaert, K. (2016). *plot3D: Plotting Multi-Dimensional Data*. Vienna, Austria: R package version 1.1.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*. 2. p.p. 16.

Speed, D. & Balding, D.J. (2014). Relatedness in the post-genomic era: is it still useful? *Nature Reviews Genetics*. 16 (1). p.pp. 33–44.

Spilman, M.S., Welbon, C., Nelson, E. & Dokland, T. (2009). Cryo-electron tomography of porcine reproductive and respiratory syndrome virus: organization of the nucleocapsid. *Journal of General Virology*. 90 (3). p.pp. 527–535.

Storgaard, T., Nielsen, H.S., Stadejek, T., Bøtner, A., Oleksiewicz, M.B. & Forsberg, R. (2001). Reversion of a live porcine reproductive and respiratory syndrome virus vaccine investigated by parallel mutations. *Journal of General Virology*. 82 (6). p.pp. 1263–1272.

Sur, J.-H., Doster, A.R., Galeota, J.A. & Osorio, F.A. (2001). Evidence for the Localization of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Antigen and RNA in Ovarian Follicles in Gilts. *Veterinary Pathology*. 38 (1). p.pp. 58–66.

Terpstra, C., Wensvoort, G. & Pol, J.M.A. (1991). Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad vims: Koch's postulates fulfilled. *Veterinary Quarterly*. 13 (3). p.pp. 131–136.

Thanawongnuwech, R., Brown, G.B., Halbur, P.G., Roth, J.A., Royer, R.L. & Thacker, B.J. (2000). Pathogenesis of Porcine Reproductive and Respiratory Syndrome Virus-induced Increase in Susceptibility to *Streptococcus suis* Infection. *Veterinary Pathology*. 37 (2). p.pp. 143–152.

The Shining (1980). *Stanley Kubrick , Stephen King, Diane Johnson*. Warner Bros. Pictures, Los Angeles, CA.

The Veterinary Record (2016). Genetic index to help breed dairy cows with greater resistance to bovine TB. *The Veterinary record*. 178 (3). p.p. 56.

Thornton, P.K. (2010). Livestock production: recent trends, future prospects. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 365 (1554). p.pp. 2853–67.

Tian, K., Yu, X., Zhao, T., Feng, Y., Cao, Z., Wang, C., Hu, Y., Chen, X., Hu, D., Tian, X., Liu, D., Zhang, S., Deng, X., Ding, Y., Yang, L., Zhang, Y., Xiao, H., Qiao, M., Wang, B., Hou, L., Wang, X., Yang, X., Kang, L., Sun, M., Jin, P., Wang, S., Kitamura, Y., Yan, J. & Gao, G.F. (2007). Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. *PLoS ONE*. 2.

Torremorell, M., Pijoan, C., Janni, K., Walker, R. & Joo, H.S. (1997). Airborne transmission of Actinobacillus pleuropneumoniae and porcine reproductive and respiratory syndrome virus in nursery pigs. *American Journal of Veterinary Research*. 58 (8). p.pp. 828–32.

Tousignant, S.J.P., Perez, A.M., Lowe, J.F., Yeske, P.E. & Morrison, R.B. (2015). Temporal and spatial dynamics of porcine reproductive and respiratory syndrome virus infection in the United States. *American Journal of Veterinary Research*. 76 (1). p.pp. 70–76.

Twine, R. (2010). *Animals as Biotechnology: Ethics, Sustainability and Critical Animal Studies*. First Edition. S. Rayner (ed.). Abingdon, UK: Routledge.

Uddin, M.J., Cinar, M.U., Grosse-Brinkhaus, C., Tesfaye, D., Tholen, E., Juengst, H., Looft, C., Wimmers, K., Phatsara, C. & Schellander, K. (2011). Mapping quantitative trait loci for innate immune response in the pig. *International Journal of Immunogenetics*. 38 (2). p.pp. 121–31.

Uddin, M.J., Grosse-Brinkhaus, C., Cinar, M.U., Jonas, E., Tesfaye, D., Tholen, E., Juengst, H., Looft, C., Ponsuksili, S., Wimmers, K., Phatsara, C. & Schellander, K. (2010). Mapping of quantitative trait loci for mycoplasma and tetanus antibodies and interferon-gamma in a porcine F(2) Duroc x Pietrain resource population. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*. 21 (7–8). p.pp. 409–18.

Uemoto, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Wilson, J.F., Rudan, I., Campbell, H., Hastie, N.D., Wright, A.F. & Haley, C.S. (2013). The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. *Frontiers in Genetics*. 4. p.p. 232.

Vannier, P. (1993). Concepts generaux sur la transmission des maladies infectieuses entre les elevages porcins et la persistance des agents infectieux au sein des elevages. *Journées de la Recherche Porcine*. 25. p.pp. 321–328.

VanRaden, P.M. (2007). Genomic Measures of Relationship and Inbreeding. *Interbull Bulletin*. (37). p.pp. 111–114.

Vazquez, A.I., Bates, D.M., Rosa, G.J.M., Gianola, D. & Weigel, K.A. (2010). Technical note: An R package for fitting generalized linear mixed models in animal breeding1. *Journal of Animal Science*. 88 (2). p.pp. 497–504.

Veerkamp, R.F., Mulder, H.A., Thompson, R. & Calus, M.P.L. (2011). Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *Journal of Dairy Science*. 94 (8). p.pp. 4189–4197.

Vidovic, V., Lukac, D., Stupar, M., Visnjic, V. & Krnjaic, J. (2012). Heritability and repeatability estimates of reproduction traits in purebred pigs. *Biotechnology in Animal Husbandry*. 28 (3). p.pp. 455–462.

Vincent, A.L., Thacker, B.J., Halbur, P.G., Rothschild, M.F. & Thacker, E.L. (2006). An investigation of susceptibility to porcine reproductive and respiratory syndrome virus between two genetically diverse commercial lines of pigs. *Journal of Animal Science*. 84 (1). p.pp. 49–57.

Vincent, A.L., Thacker, B.J., Halbur, P.G., Rothschild, M.F. & Thacker, E.L. (2005). In Vitro Susceptibility of Macrophages to Porcine Reproductive and Respiratory Syndrome Virus Varies between Genetically Diverse Lines of Pigs. *Viral Immunology*. 18 (3). p.pp. 506–512.

Vinkhuyzen, A.A.E., Wray, N.R., Yang, J., Goddard, M.E. & Visscher, P.M. (2013). Estimation and Partition of Heritability in Human Populations Using Whole-Genome Analysis Methods. *Annual Review of Genetics*. 47 (1). p.pp. 75–95.

Voicu, I.L., Silim, A., Morin, M. & Elazhary, M.A. (1994). Interaction of porcine reproductive and respiratory syndrome virus with swine monocytes. *The Veterinary record*. 134 (16). p.pp. 422–3.

Vonnahme, K.A., Wilson, M.E., Foxcroft, G.R. & Ford, S.P. (2002). Impacts on conceptus survival in a commercial swine herd. *Journal of animal science*. 80 (3). p.pp. 553–9.

Vu, H.L.X., Kwon, B., Yoon, K.-J., Laegreid, W.W., Pattnaik, A.K. & Osorio, F.A. (2011). Immune Evasion of Porcine Reproductive and Respiratory Syndrome Virus through Glycan Shielding Involves both Glycoprotein 5 as Well as Glycoprotein 3. *Journal of Virology*. 85 (11). p.pp. 5555–5564.

van der Waaij, E.H., Hazeleger, W., Soede, N.M., Laurenssen, B.F.A. & Kemp, B. (2010). Effect of excessive, hormonally induced intrauterine crowding in the gilt on fetal development on day 40 of pregnancy. *Journal of Animal Science*. 88 (8). p.pp. 2611–2619.

Warnes, G.R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J. & Et, A.. (2015). *gdata: Various R Programming Tools for Data Manipulation*.

Weir, B.S. (1979). Inferences about Linkage Disequilibrium. *Biometrics*. 35 (1). p.p. 235.

Wells, K.D., Bardot, R., Whitworth, K.M., Trible, B.R., Fang, Y., Mileham, A., Kerrigan, M.A., Samuel, M.S., Prather, R.S. & Rowland, R.R.R. (2017). Replacement of Porcine CD163 Scavenger Receptor Cysteine-Rich Domain 5 with a CD163-Like Homolog Confers Resistance of Pigs to Genotype 1 but Not Genotype 2 Porcine Reproductive and Respiratory Syndrome Virus. *Journal of Virology*. 91 (2). p.pp. e01521-16.

Wensvoort, G., Terpstra, C., Pol, J.M.A., ter Laak, E.A., Bloemraad, M., de Kluyver, E.P., Kragten, C., van Buiten, L., den Besten, A., Wagenaar, F., Broekhuijsen, J.M., Moonen, P.L.J.M., Zetstra, T., de Boer, E.., Tibben, H.J., de Jong, M.F., van 't Veld, P., Greenland, G.J.R., van Gennep, J.A., Voets, M.T., Verheijden, J.H.M. & Braamskamp, J. (1991). Mystery swine disease in the Netherlands: The isolation of Lelystad virus. *Veterinary Quarterly*. 13 (3). p.pp. 121–130.

Whiting, T.L. & Pasma, T. (2008). Isolated weaning technology: humane benefits and concerns in the production of pork. *The Canadian Veterinary Journal*. 49 (3). p.pp. 293–301.

Whitworth, K.M. & Prather, R.S. (2017). Gene editing as applied to prevention of reproductive porcine reproductive and respiratory syndrome. *Molecular Reproduction and Development*. 9999. p.pp. 1–8.

Whitworth, K.M., Rowland, R.R.R., Ewen, C.L., Trible, B.R., Kerrigan, M.A., Cino-Ozuna, A.G., Samuel, M.S., Lightner, J.E., McLaren, D.G., Mileham, A.J., Wells, K.D. & Prather, R.S. (2015). Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nature Biotechnology*. 34 (1). p.pp. 20–22.

Wills, R.W., Zimmerman, J.J., Swenson, S.L., Yoon, K.-J., Hill, H.T., Bundy, D.S. & McGinley, M.J. (1997). Transmission of PRRSV by direct, close, or indirect contact. *Journal of Swine Health and Production*. 5 (6). p.pp. 213–218.

Winters, M. (2016). *Breeding for TB resistance – the 'TB advantage'*. Buxton Derbyshire.

Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*. 56 (645). p.pp. 330–338.

Wright, S. (1921). Systems of Mating. I. the Biometric Relations between Parent and Offspring. *Genetics*. 6 (2). p.pp. 111–23.

Yaeger, M., Prieve, T., Collins, J., Christopher-Hennings, J., Nelson, E. & Benfield, D. (1993). Evidence for transmission of porcine reproductive and respiratory syndrome (PRRS) virus in boar semen. *Swine Health and Production*. 1 (5). p.pp. 7–9.

Yagi, R., Zhu, J. & Paul, W.E. (2011). An updated view on transcription factor GATA3-mediated regulation of Th1 and Th2 cell differentiation. *International immunology*. 23 (7). p.pp. 415–20.

Yang, T., Wilkinson, J., Wang, Z., Ladinig, A., Harding, J. & Plastow, G. (2016). A genome-wide association study of fetal response to type 2 porcine reproductive and respiratory syndrome virus challenge. *Scientific Reports*. 6 (1). p.p. 20305.

Yoo, D., Song, C., Sun, Y., Du, Y., Kim, O. & Liu, H.-C. (2010). Modulation of host cell responses and evasion strategies for porcine reproductive and respiratory syndrome virus. *Virus Research*. 154 (1–2). p.pp. 48–60.

Yoon, I.J., Joo, H.S., Christianson, W.T., Morrison, R.B. & Dial G D (1993). Persistent and contact infection in nursery pigs experimentally infected with porcine reproductive and respiratory syndrome (PRRS) virus. *Swine Health and Production*. 1 (4). p.pp. 5–8.

Zeggini, E. & Morris, A. (2011). *Analysis of Complex Disease Association Studies : A Practical Guide*. Elsevier.

Zhang, Y.-Z., Zhou, D.-J., Qin, X.-C., Tian, J.-H., Xiong, Y., Wang, J.-B., Chen, X.-P., Gao, D.-Y., He, Y.-W., Jin, D., Sun, Q., Guo, W.-P., Wang, W., Yu, B., Li, J., Dai, Y.-A., Li, W., Peng, J.-S., Zhang, G.-B., Zhang, S., Chen, X.-M., Wang, Y., Li, M.-H., Lu, X., Ye, C., de Jong, M.D. & Xu, J. (2012). The Ecology, Genetic Diversity, and Phylogeny of Huaiyangshan Virus in China. *Journal of Virology*. 86 (5). p.pp. 2864–2868.

Zheng, G., Freidlin, B. & Gastwirth, J.L. (2006). Robust genomic control for association studies. *American journal of human genetics*. 78 (2). p.pp. 350–6.

Zhuang, Q., Barfod, K., Wachmann, H., Mortensen, S. & Willeberg, P. (2002). Serological surveillance for PRRS in Danish genetic pig herds and risk factors for PRRS infection. In: *Proceedings International Pig Veterinary Society*. 2002, Ames, Iowa, p. 2:231.

Zimmerman, J., Yoon, K.-J. & Neumann, E. (2003). *2003 PRRS Compendium Producer Edition: A Reference for Pork Producers*. Des Moines: National Pork Board.

Zondervan, K.T. (2011). *Analysis of Complex Disease Association Studies*. E. Zeggini (ed.). Amsterdam, Netherlands: Elsevier.

Zuckermann, F.A., Garcia, E.A., Luque, I.D., Christopher-Hennings, J., Doster, A., Brito, M. & Osorio, F. (2007). Assessment of the efficacy of commercial porcine reproductive and respiratory syndrome virus (PRRSV) vaccines based on measurement of serologic response, frequency of gamma-IFN-producing cells and virological parameters of protection upon challenge. *Veterinary Microbiology*. 123 (1). p.pp. 69–85.

Zumbach, B., Misztal, I., Tsuruta, S., Holl, J., Herring, W. & Long, T. (2007). Genetic correlations between two strains of Durocs and crossbreds from differing production environments for slaughter traits. *Journal of Animal Science*. 85 (4). p.p. 901.

Appendix

Additional Figures

This section contains additional figures produced not deemed sufficiently informative to be included in the main body.

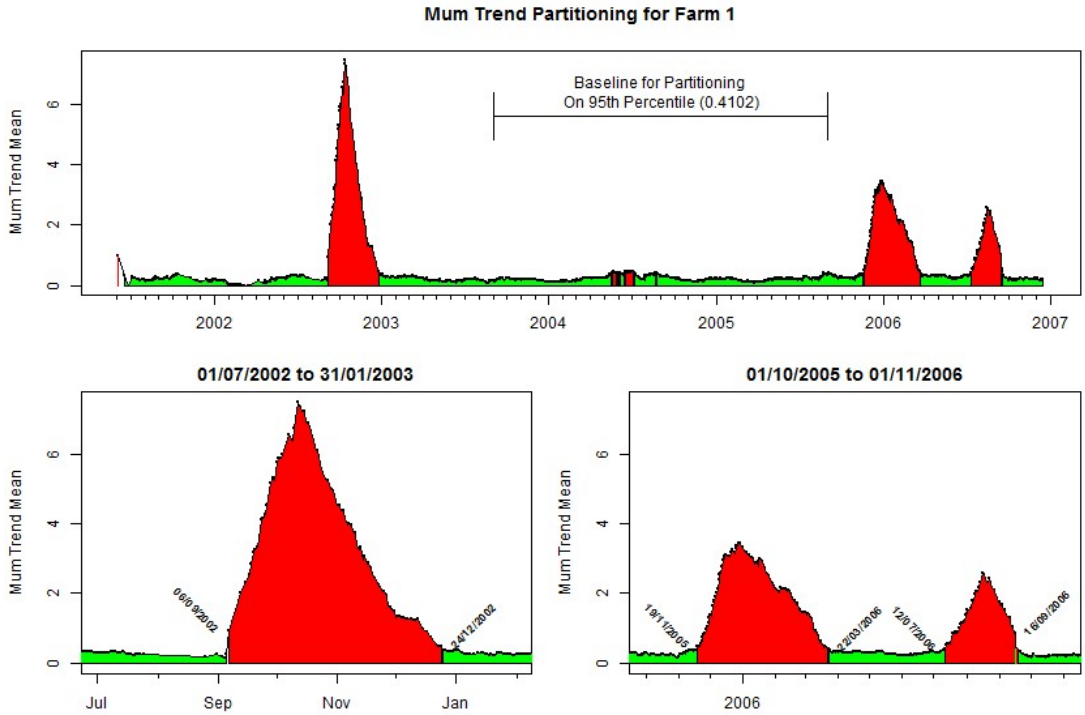

Figure A.1 – Mum Trend Partitioning for Farm 1

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) for and below the threshold (green)

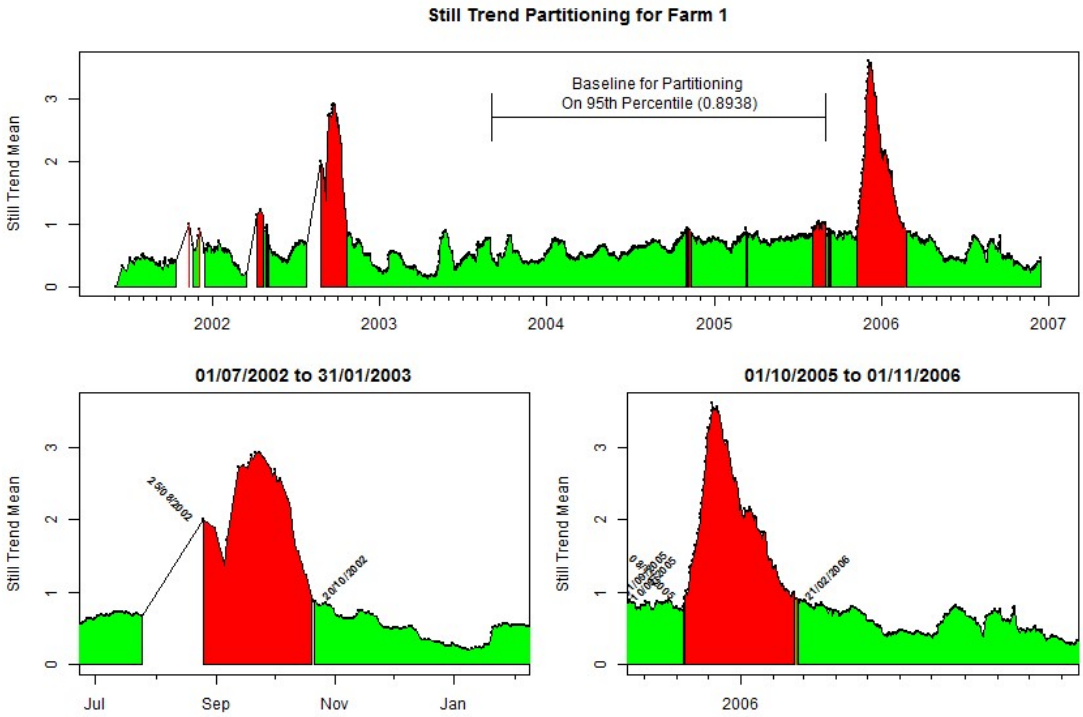

Figure A.2 – Still Trend Partitioning for Farm 1

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) and below the threshold (green)

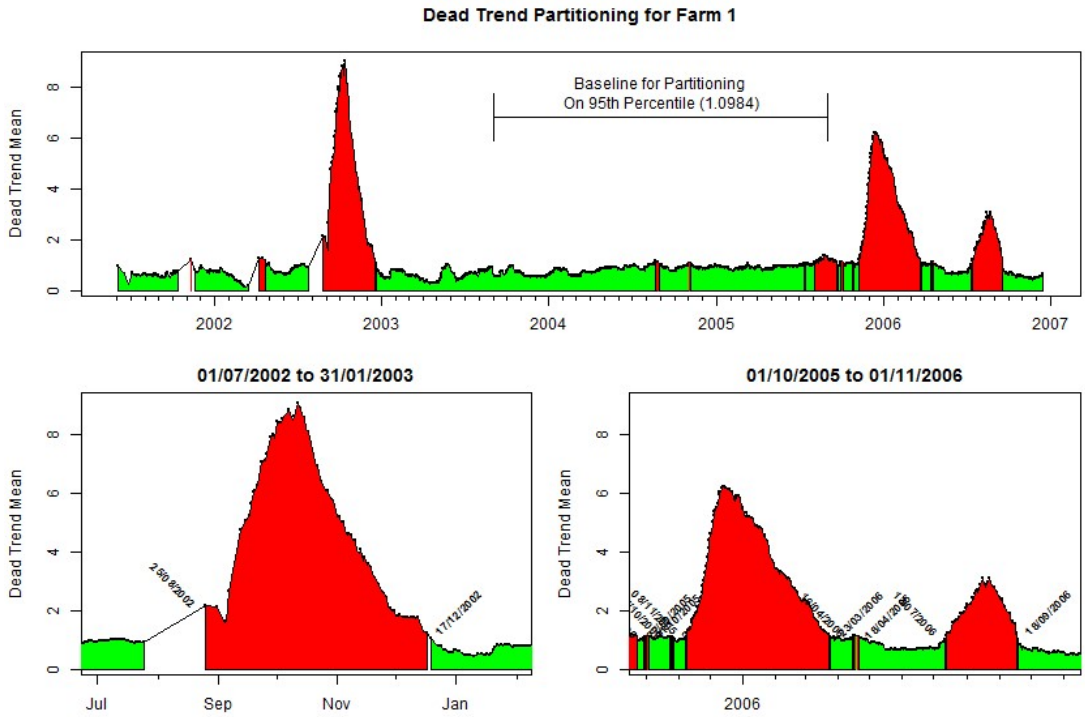

Figure A.3 – Dead Trend Partitioning for Farm 1

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) and below the threshold (green)

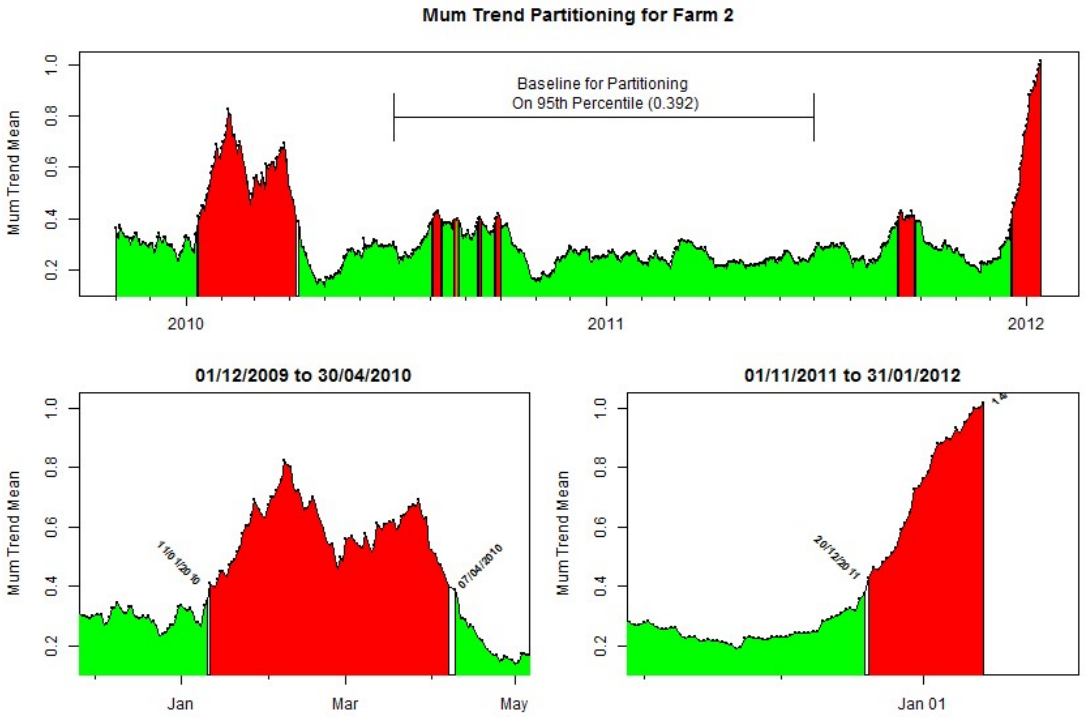

Figure A.4 – Mum Trend Partitioning for Farm 2

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) and below the threshold (green)

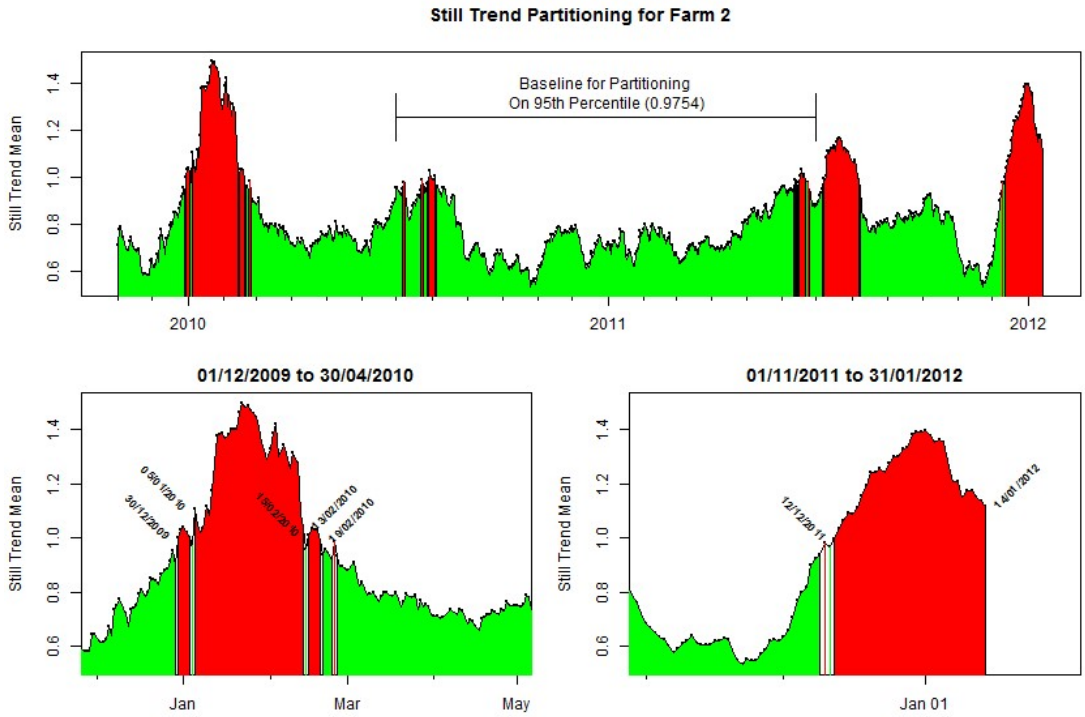

Figure A.5 – Still Trend Partitioning for Farm 2

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) and below the threshold (green)

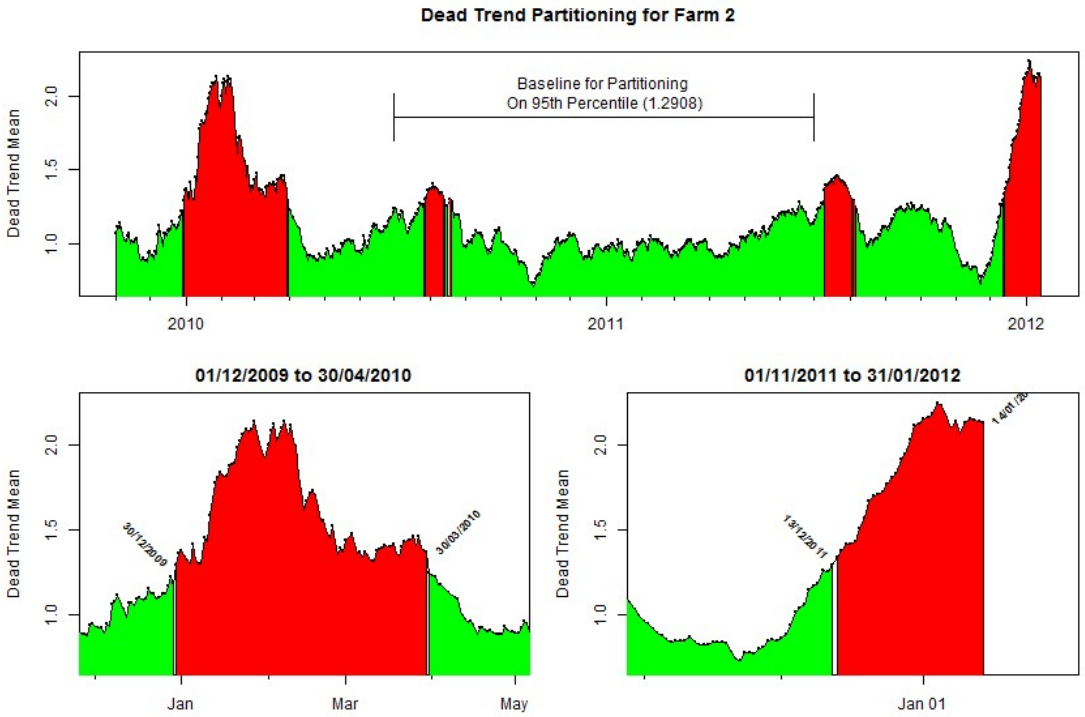

Figure A.6 – Dead Trend Partitioning for Farm 2

Rolling 30-day trait trend identifying periods where the trend is over the baseline threshold (red) and below the threshold (green)

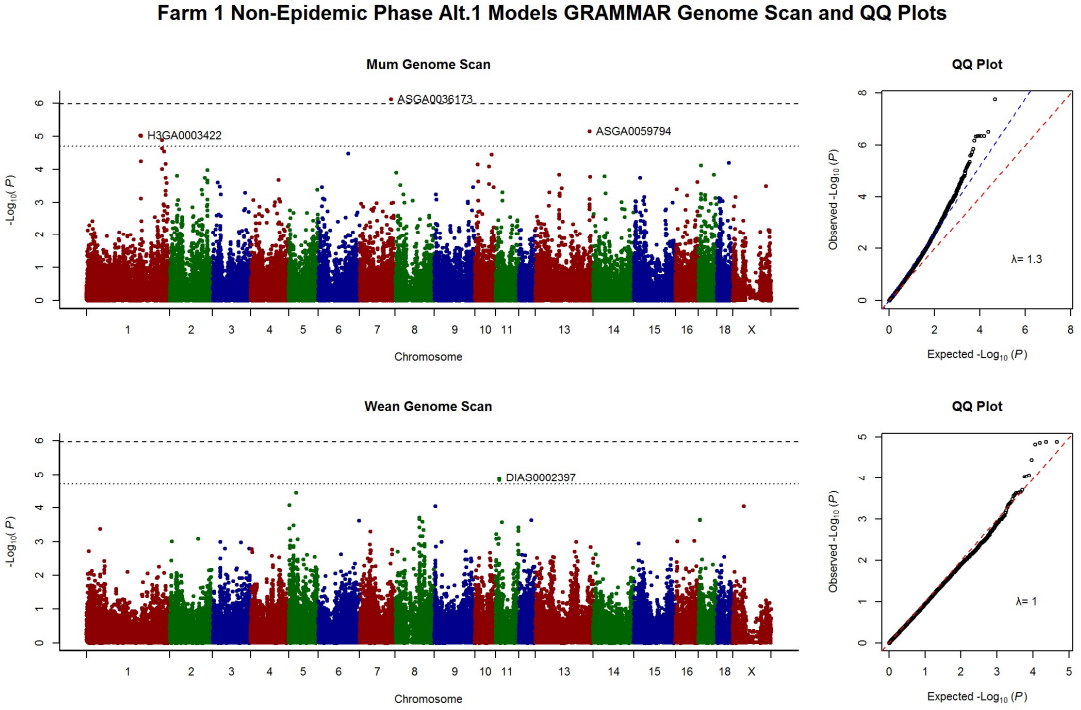

Figure A.7 – Farm 1 Manhattan Plots using GRAMMAR for Mum and Wean Traits Using the Basic Model with Non-Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,040 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line)

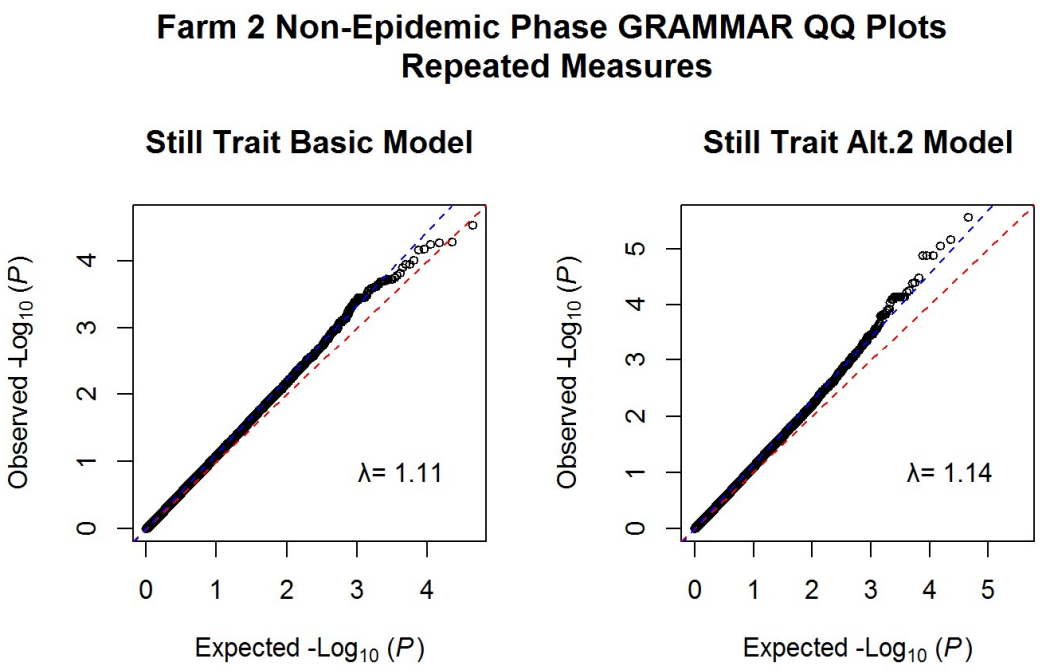

Figure A.8 – Farm 2 Non-Epidemic Phase GRAMMAR QQ Plots

Expected and observed -$\log_{10}$(P) values from the GRAMMAR score applied to the Farm 2 Epidemic phase analysis using repeated measures basic and Alt.2 models. QQ plots show x=y (red dashed line) degree of inflation indicated (blue dashed line)

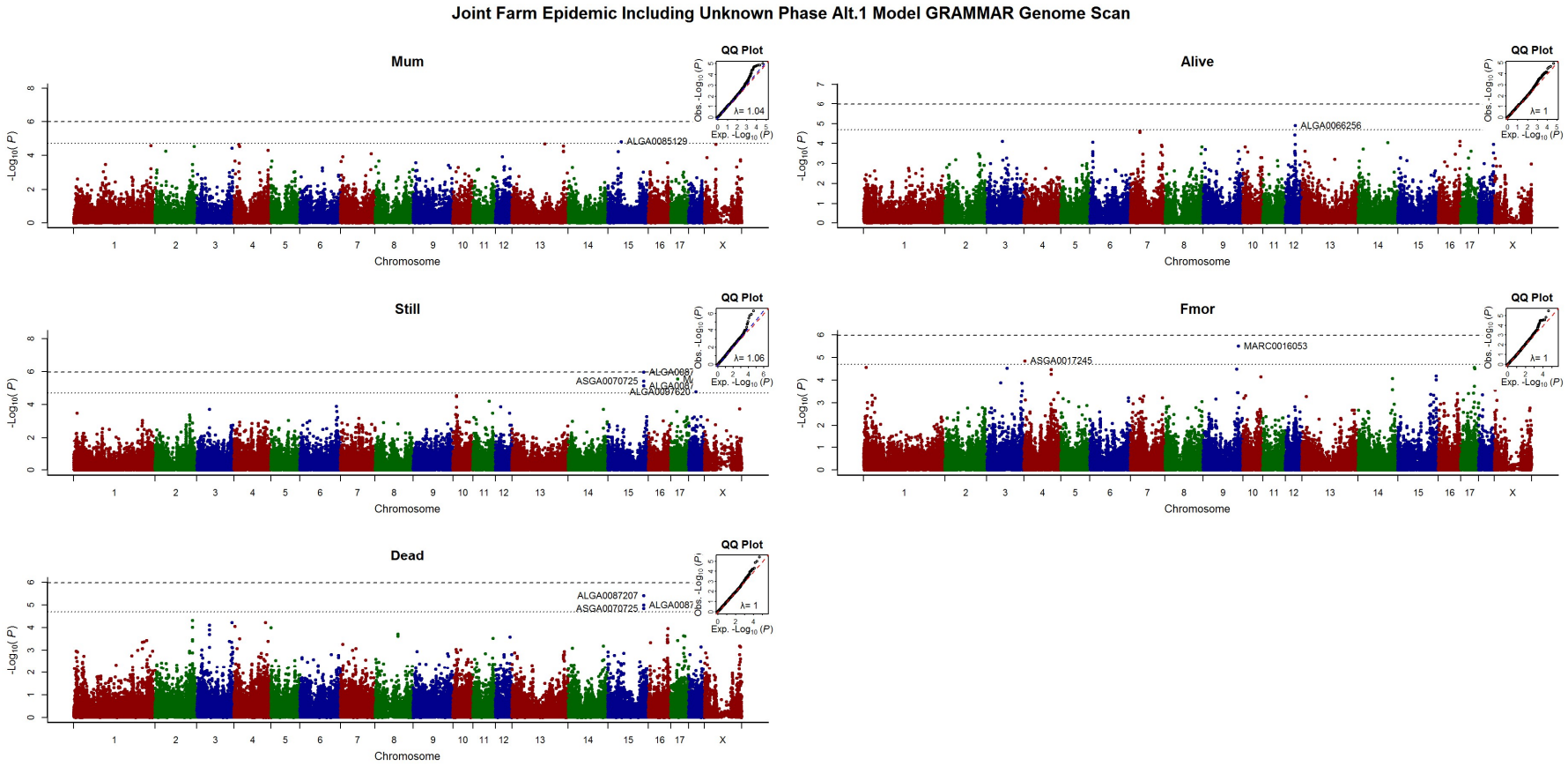

Figure A.9 – Joint Farm Manhattan Plots using GRAMMAR for All Disease Indicator Traits Using the Basic Model with Epidemic Including Unknown Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 48,093 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 degree of inflation indicated (blue dashed line).

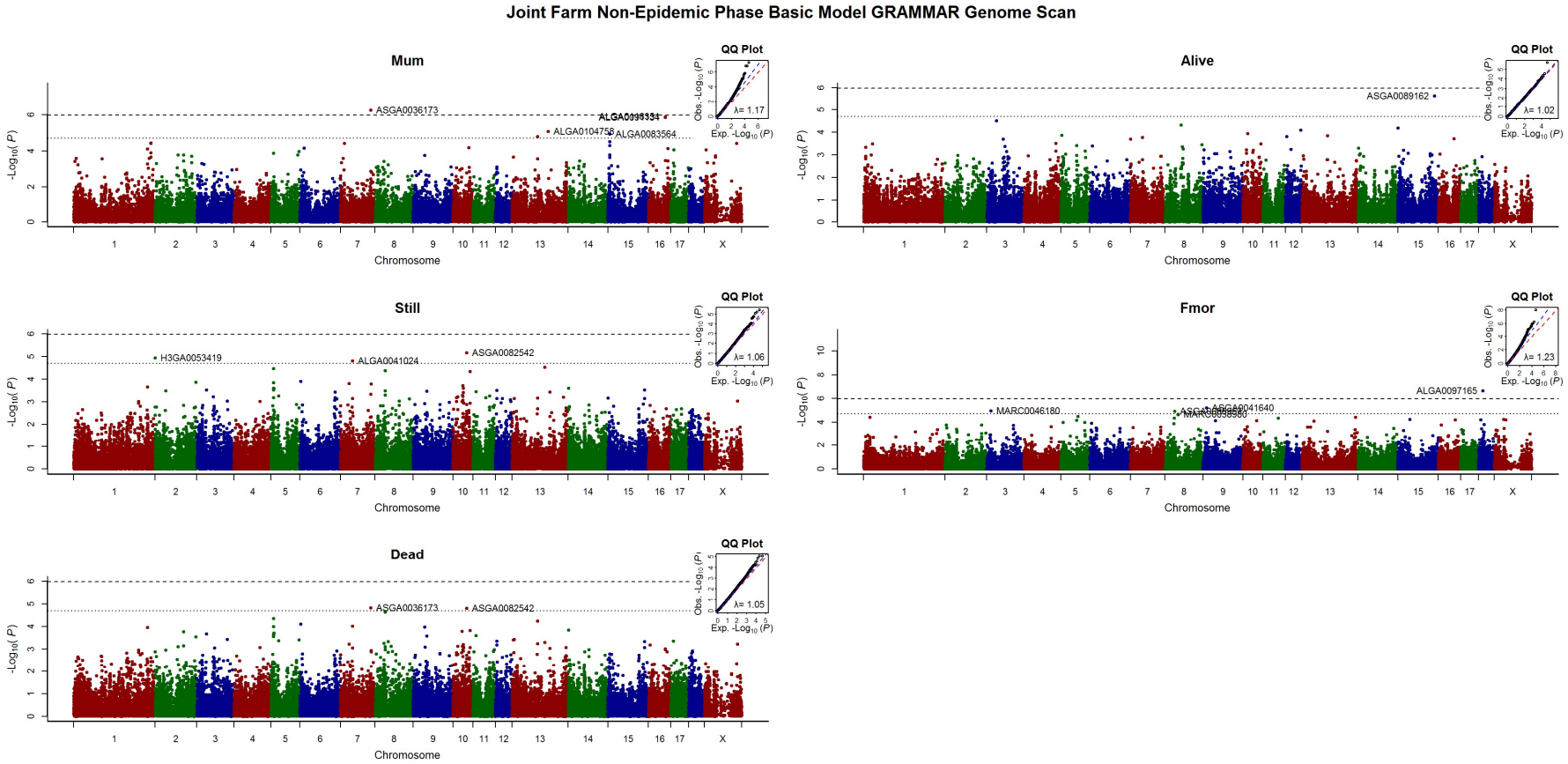

Figure A.10 – Joint Farm Manhattan Plots using GRAMMAR for All Disease Indicator Traits Using the Basic Model with Non-Epidemic Phase Data

Bonferroni corrected significance thresholds shown at the genome wide (dashed line) and chromosome level (dotted line), based on 47,941 SNPs. SNPs significant above the indicative level shown labelled. QQ plots show x=y (red dashed line) where λ>1 inflation indicated (blue dashed line)

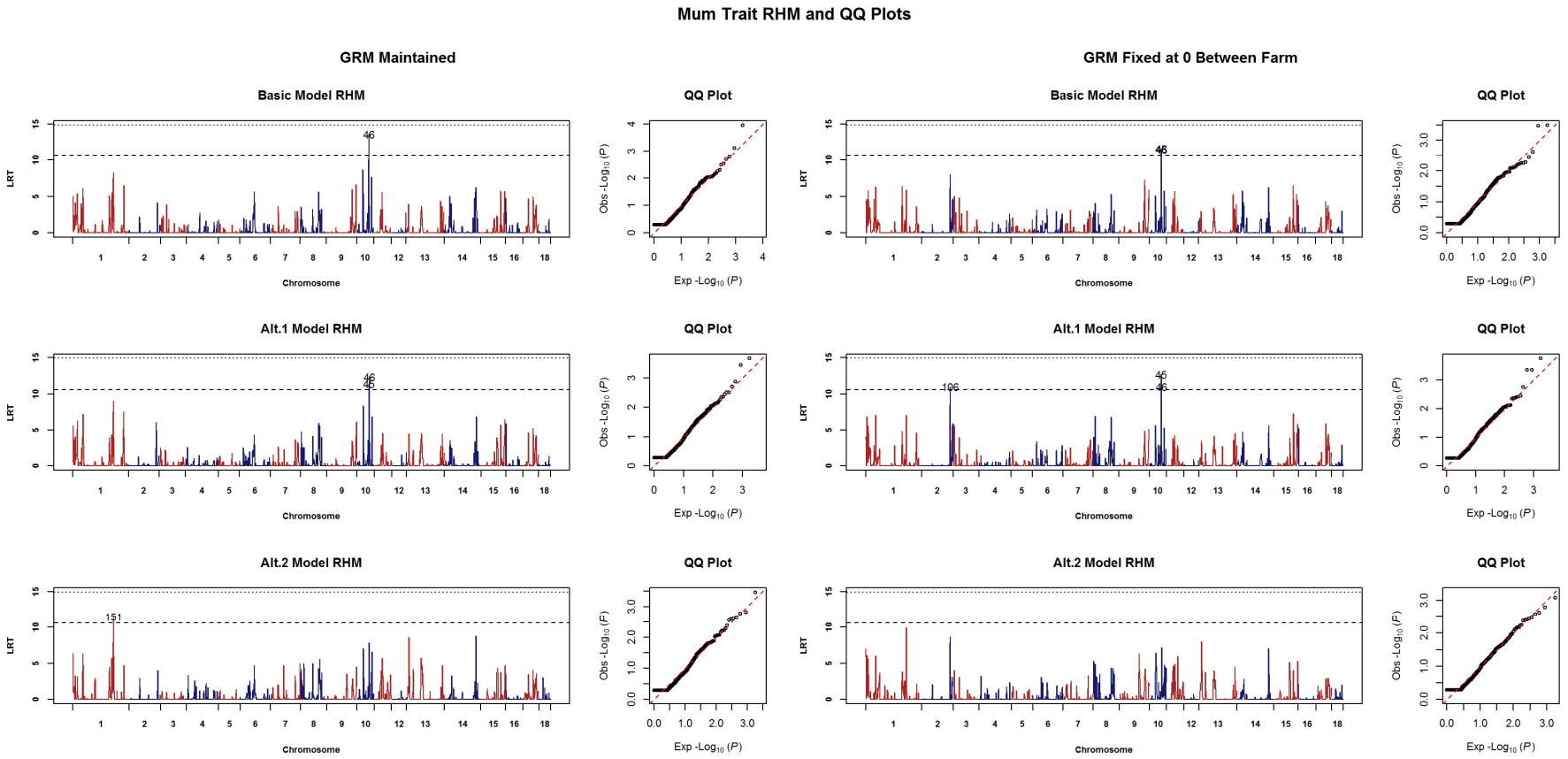

Figure A.11 – Mum Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Mum* trait. All three models shown, GRMs used are whole population maintained (left) and between farm fixed at Zero (right). Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values and plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

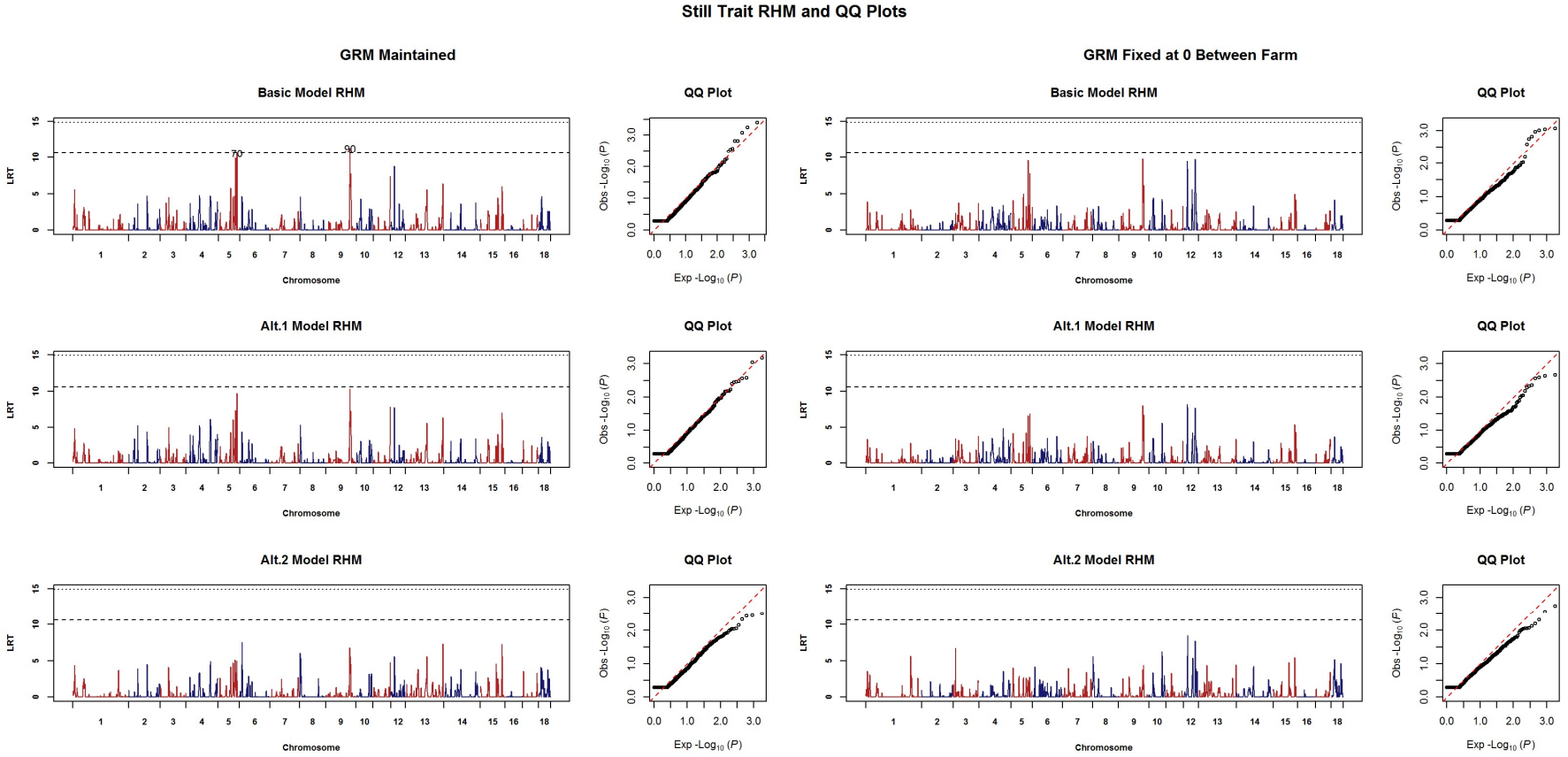

Figure A.12 – Still Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Still* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -log10(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

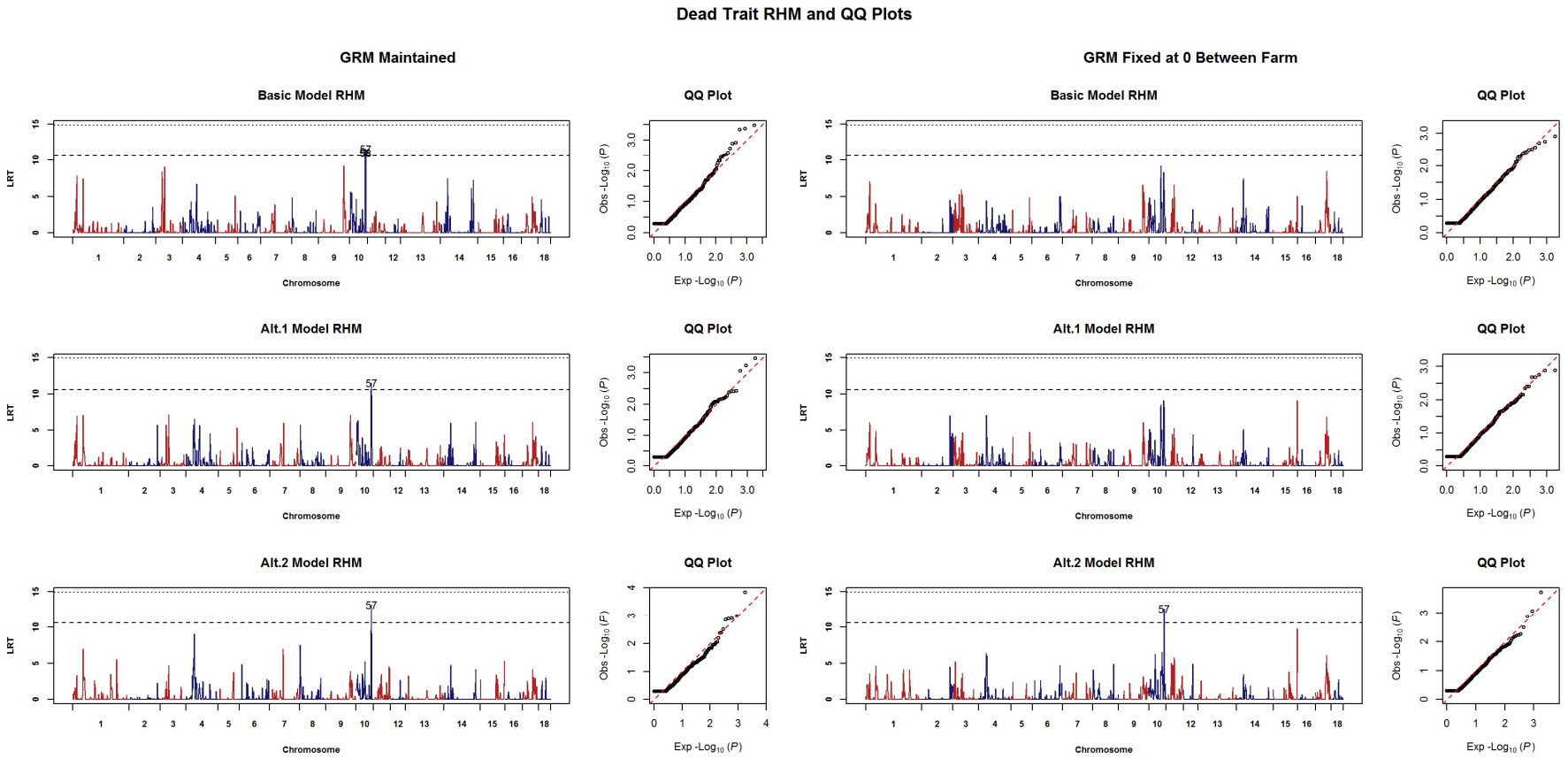

Figure A.13 – Dead Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Dead* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -log10(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

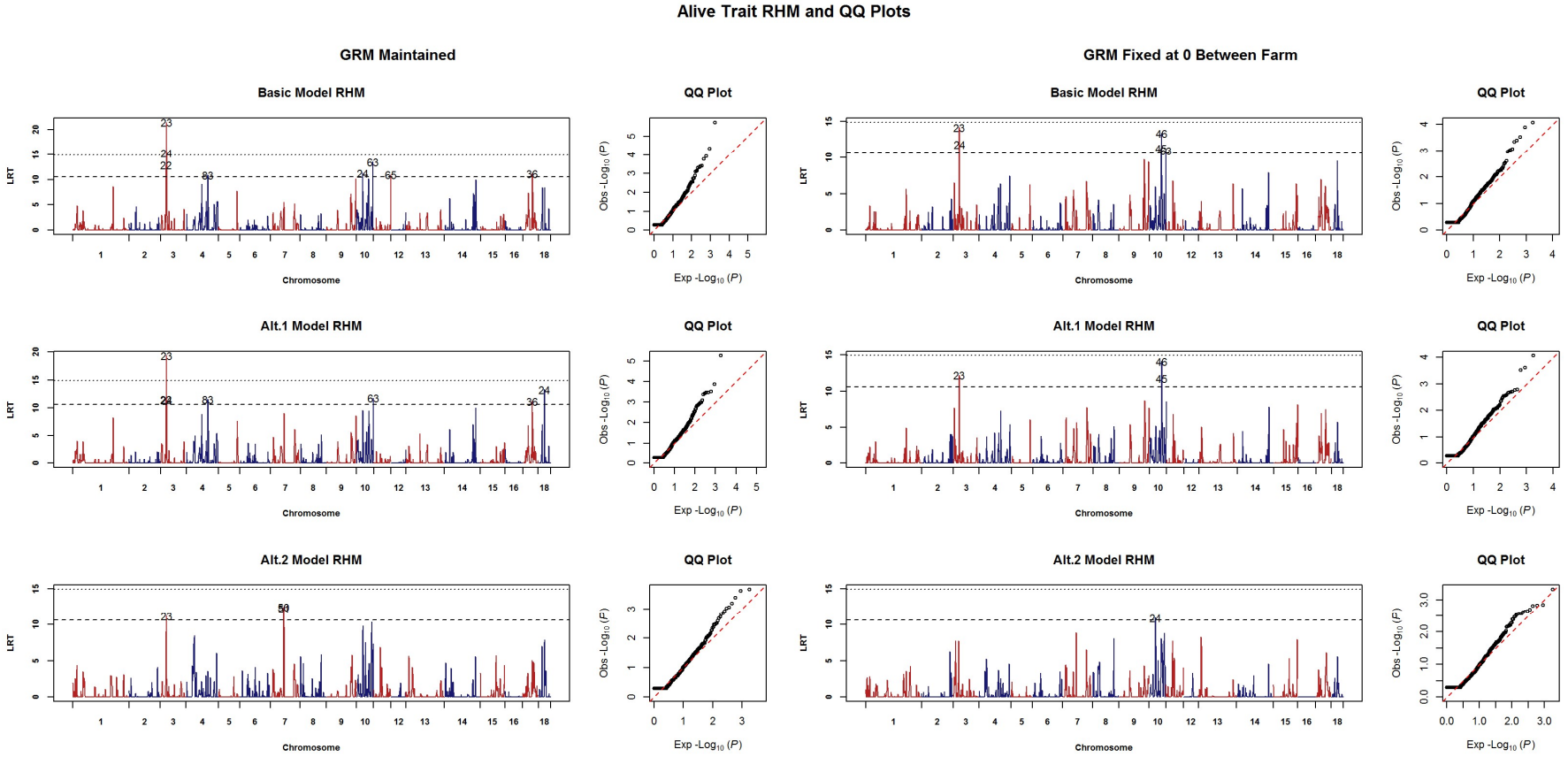

Figure A.14 – Alive Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Alive* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -log10(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

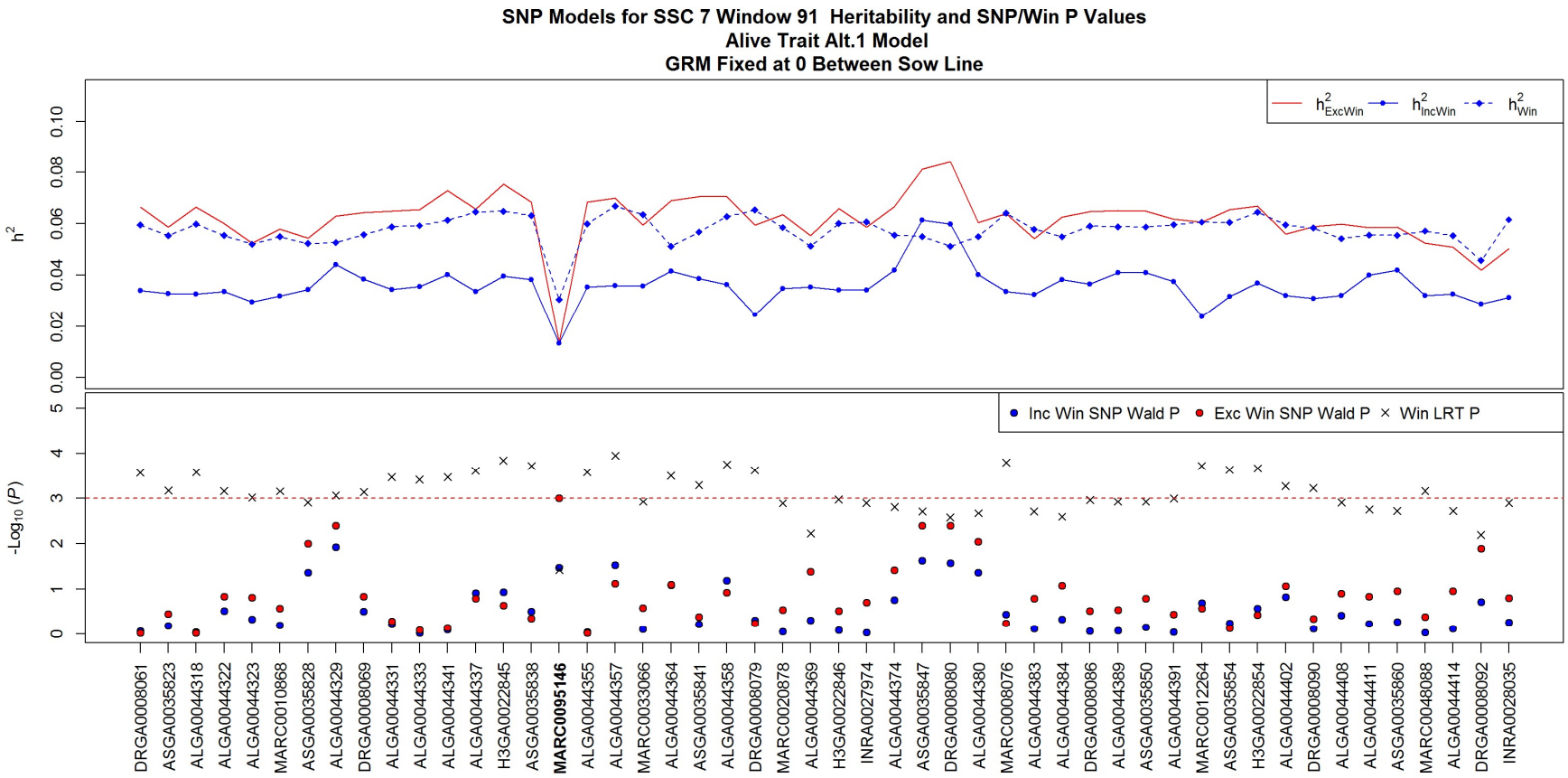

Figure A.15 – Alive Trait, Alt.1 Model, GRM Between Sow Line Fixed at Zero SNPs Analysis for SSC7 Window 91, Heritability and SNP/Window P-values

SNP and Window estimates using the reduced model (red) the full model (blue). (Top) h^2 (solid lines) and h^2_{Win} (dashed line). (Bottom) -$\log_{10}$(P) values shown using the Wald F statistic for SNPs and LRT for the window. Bonferroni corrected P<0.05 value for 50 tests (red dashed line).

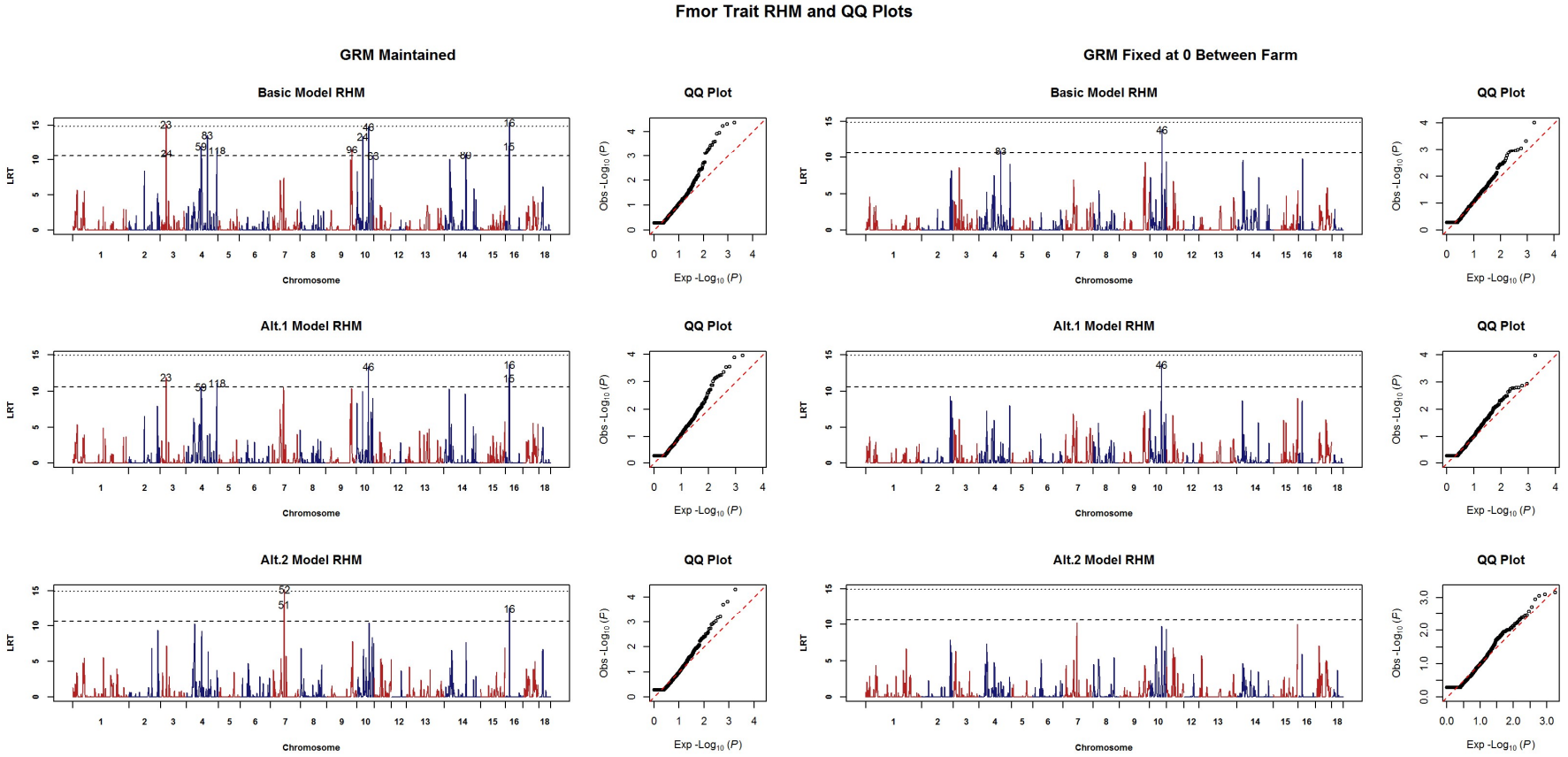

Figure A.16 – Fmor Trait Regional Heritability Mapping Window LRT Statistic by Chromosome Window Position

Likelihood ratio test statistic of each 50 SNP window at 25 SNP intervals, for *Fmor* trait. All three models shown, between sow line GRM fixed at Zero. Significance thresholds shown at the genome wide level (dotted line) and indicative level (dashed line). QQ plots show the observed, ranked -$\log_{10}$(P) values plotted against those expected under a null distribution. The line of unity (Y=X) shown indicated (red dashed line).

Additional Tables

Table A.1 – Source of the Reported Traits

Trait	Description	Source
Mum	Number of mummified Piglets per Litter	Raw Data
Still	Number of Stillborn Piglets per Litter	Raw Data
Dead	Number of Dead Piglets per Litter	Mum+Still
Alive	Number of Alive Piglets per Litter	Raw Data
Tof	Number of Total observable foetuses per litter	Mum+Still+Alive
Fmor	Farrowing Mortality, proportion of dead piglets per litter expressed as a fraction of the total observable foetuses per litter	$\frac{Mum + Still}{Mum + Still + Alive}$
Gest	Gestation length, the farrowing date minus the service date in days	Farrow Date– Service Date (Days)

Table A.2 – Farm 2 Including Unknown Epidemic F Statistic P-values for All Fixed Effects, Covariates and Interactions under the Basic and Alternative Models for All Traits

Model	Trait	Parity.	Sow Line.	Tof	Epidemic†	Epidemic† × Trend‡	Epidemic × Parity.	Epidemic† × Sow Line.
Basic	Mum¥	0.163	0.095	$<3\times10^{-16}$	-	-	-	-
Basic	Still¥	$<3\times10^{-16}$	7×10^{-7}	$<3\times10^{-16}$	-	-	-	-
Basic	Dead¥	3×10^{-16}	1×10^{-5}	$<3\times10^{-16}$	-	-	-	-
Basic	Alive	$<3\times10^{-16}$	0.101	$<3\times10^{-16}$	-	-	-	-
Basic	Tof	5×10^{-12}	1×10^{-10}	-	-	-	-	-
Basic	Fmor	$<3\times10^{-16}$	7×10^{-7}	-	-	-	-	-
Basic	Gest	$<3\times10^{-16}$	-	$<3\times10^{-16}$	-	-	-	-
Alt. 1	Mum¥	0.011	0.236	$<3\times10^{-16}$	1×10^{-8}	-	1×10^{-4}	-
Alt. 1	Still¥	$<3\times10^{-16}$	1×10^{-6}	$<3\times10^{-16}$	1×10^{-11}	-	-	-
Alt. 1	Dead¥	3×10^{-4}	4×10^{-5}	$<3\times10^{-16}$	2×10^{-10}	-	0.021	-
Alt. 1	Alive	$<3\times10^{-16}$	0.204	$<3\times10^{-16}$	1×10^{-6}	-	-	0.022
Alt. 1	Tof	1×10^{-7}	3×10^{-6}	-	0.707	-	-	0.045
Alt. 1	Fmor	0.002	3×10^{-6}	-	$<3\times10^{-16}$	-	6×10^{-6}	-
Alt. 1	Gest	$<3\times10^{-16}$	-	$<3\times10^{-16}$	2×10^{-7}	-	-	-
Alt. 2	Mum¥	0.007	0.146	$<3\times10^{-16}$	0.042	$<3\times10^{-16}$	1×10^{-4}	-
Alt. 2	Still¥	$<3\times10^{-16}$	3×10^{-6}	$<3\times10^{-16}$	0.843	$<3\times10^{-16}$	-	-
Alt. 2	Dead¥	1×10^{-4}	6×10^{-5}	$<3\times10^{-16}$	0.335	$<3\times10^{-16}$	0.028	-
Alt. 2	Alive	$<3\times10^{-16}$	0.614	$<3\times10^{-16}$	0.183	$<3\times10^{-16}$	-	0.026
Alt. 2	Tof	1×10^{-7}	3×10^{-5}	-	0.366	6×10^{-5}	-	0.053
Alt. 2	Fmor	0.002	1×10^{-5}	-	0.085	$<3\times10^{-16}$	1×10^{-5}	-
Alt. 2	Gest	$<3\times10^{-16}$	-	$<3\times10^{-16}$	0.001	$<3\times10^{-16}$	-	-

¥Trait transformed by log(trait+1)

†When fitted in the basic model this effect is the Phase term, when fitted in the alternative models this is the Epidemic ID term

‡The fitted trend is the 30 day rolling mean corresponding to the trait for *Mum, Still, Dead*. for all other traits the *Dead* Trend was used

Table A.3 – Estimates of Model Fit, Variance Components and Variance Component Estimate Significance for Farm 2 Models

Trait	Model	Akaike Information Criterion	Log Likelihood	Root Mean Square Error	Sow Log LRT P
Mum	Basic	4826.938	-2390.469	0.342	5×10^{-4}
Mum	Alt. 1	4844.081	-2391.04	0.343	0.001
Mum	Alt. 2	4751.659	-2341.83	0.339	0.001
Still	Basic	8414.785	-4191.392	0.443	3×10^{-16}
Still	Alt. 1	8420.856	-4193.428	0.443	3×10^{-16}
Still	Alt. 2	8356.854	-4158.427	0.439	$<3\times10^{-16}$
Dead	Basic	9410.965	-4682.482	0.484	7×10^{-12}
Dead	Alt. 1	9429.652	-4683.826	0.484	8×10^{-12}
Dead	Alt. 2	9359.981	-4645.991	0.479	7×10^{-13}
Alive	Basic	20959.967	-10459.983	1.262	4×10^{-11}
Alive	Alt. 1	20965.654	-10458.827	1.262	4×10^{-11}
Alive	Alt. 2	20872.027	-10409.014	1.246	3×10^{-12}
Tof	Basic	30532.674	-15251.337	2.518	$<3\times10^{-16}$
Tof	Alt. 1	30514.239	-15234.119	2.514	$<3\times10^{-16}$
Tof	Alt. 2	30497.856	-15222.928	2.511	$<3\times10^{-16}$
Fmor	Basic	-10588.212	5316.106	0.094	1×10^{-9}
Fmor	Alt. 1	-10550.244	5305.122	0.093	1×10^{-9}
Fmor	Alt. 2	-10630.775	5348.387	0.092	6×10^{-11}
Wean	Basic	25584.995	-12753.498	1.81	2×10^{-15}
Wean	Alt. 1	25335.592	-12620.796	1.772	6×10^{-16}
Wean	Alt. 2	25237.077	-12568.539	1.75	$<3\times10^{-16}$
Gest	Basic	19171.881	-9573.941	0.837	$<3\times10^{-16}$
Gest	Alt. 1	19153.519	-9563.76	0.835	$<3\times10^{-16}$
Gest	Alt. 2	19058.975	-9513.487	0.824	$<3\times10^{-16}$

¥Transformed log(trait + 1)

Table A.4 – Least Square Means for Epidemic Factor Across Models All Traits Farm 2 Data Including Unknown Epidemic

Model	Epidemic Factor	Mum¥	Still¥	Dead¥	Alive	Tof	Fmor	Gest
Basic	Non-Epidemic	0.209 (0.016)	0.613 (0.03)	0.854 (0.037)	12.521 (0.052)	13.485 (0.135)	0.079 (0.004)	115.866 (0.029)
Basic	Epidemic	0.362 (0.024)	0.85 (0.046)	1.278 (0.061)	11.832 (0.069)	13.486 (0.167)	0.132 (0.005)	115.748 (0.049)
Alt. 1	Non-Epidemic†	0.209 (0.016)	0.614 (0.03)	0.855 (0.037)	12.522 (0.052)	13.473 (0.134)	0.079 (0.004)	115.83 (0.03)
Alt. 1	Epidemic 4	0.353 (0.031)	0.823 (0.057)	1.25 (0.076)	11.792 (0.088)	13.896 (0.204)	0.134 (0.006)	115.934 (0.062)
Alt. 1	Epidemic 5	0.371 (0.031)	0.878 (0.059)	1.307 (0.078)	11.873 (0.088)	13.073 (0.205)	0.129 (0.006)	115.512 (0.069)
Alt. 2‡	Non-Epidemic†	0.189 (0.013)	0.486 (0.019)	0.623 (0.02)	12.504 (0.052)	13.482 (0.134)	0.081 (0.004)	115.848 (0.03)
Alt. 2‡	Epidemic 4	0.301 (0.023)	0.607 (0.031)	0.816 (0.034)	11.776 (0.087)	13.904 (0.204)	0.135 (0.006)	115.937 (0.061)
Alt. 2‡	Epidemic 5	0.314 (0.023)	0.638 (0.031)	0.842 (0.034)	11.853 (0.087)	13.083 (0.205)	0.131 (0.006)	115.531 (0.069)

¥Traits transformed by log(trait+1), least square mean values are back-transformed onto the response scale.
‡ Estimates are corrected for Epidemic ID×Trend interaction by fixing covariate estimate at the mean for each epidemic.

Table A.5 – Farm 1 Variance Components and Heritability Estimated Using the A Matrix with Single Record Data for Non-Epidemic Phase and Epidemic Phase

		Non- Epidemic Phase					-Epidemic phase				
Trait	Model	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum[¥]	Basic	-	0.11 (0.01)	0.11 (0.01)	-	-	0.04 (0.04)	0.65 (0.05)	0.7 (0.04)	0.06 (0.06)	0.09
Still[¥]	Basic	2×10^{-3} (0.01)	0.18 (0.01)	0.18 (0.01)	0.01 (0.05)	0.6	0.08 (0.04)	0.38 (0.03)	0.46 (0.03)	0.16 (0.07)	<0.001
Dead[¥]	Basic	-	0.25 (0.01)	0.25 (0.01)	-	-	0.05 (0.04)	0.68 (0.05)	0.72 (0.04)	0.07 (0.06)	0.08
Alive	Basic	0.01 (0.07)	1.52 (0.1)	1.53 (0.08)	0.01 (0.05)	0.62	1.26 (0.93)	12.82 (1.04)	14.09 (0.87)	0.09 (0.06)	0.03
Tof	Basic	1.53 (0.57)	6.57 (0.53)	8.11 (0.48)	0.19 (0.07)	<0.001	0.93 (0.6)	7.73 (0.64)	8.66 (0.54)	0.11 (0.07)	0.02
Fmor	Basic	-	0.01 (8×10^{-4})	0.01 (8×10^{-4})	-	-	0.01 (0.01)	0.11 (0.01)	0.12 (0.01)	0.1 (0.06)	0.02
Wean	Basic	-	2.3 (0.12)	2.3 (0.12)	-	-	2.98 (0.88)	6.81 (0.67)	9.79 (0.65)	0.3 (0.08)	<0.001
Gest	Basic	0.49 (0.17)	1.93 (0.15)	2.42 (0.14)	0.2 (0.06)	<0.001	1.5 (0.53)	4.37 (0.43)	5.87 (0.39)	0.26 (0.08)	<0.001
Mum[¥]	Alt.1[†]						0.03 (0.04)	0.65 (0.05)	0.68 (0.04)	0.05 (0.06)	0.15
Still[¥]	Alt.1[†]						0.05 (0.03)	0.39 (0.03)	0.44 (0.03)	0.1 (0.07)	0.03
Dead[¥]	Alt.1[†]						0.01 (0.04)	0.67 (0.05)	0.68 (0.04)	0.02 (0.05)	0.42
Alive	Alt.1[†]						0.57 (0.82)	12.97 (1.01)	13.54 (0.83)	0.04 (0.06)	0.21
Tof	Alt.1[†]						0.72 (0.57)	7.86 (0.64)	8.57 (0.53)	0.08 (0.06)	0.05
Fmor	Alt.1[†]						0.01 (0.01)	0.11 (0.01)	0.12 (0.01)	0.05 (0.06)	0.17
Wean	Alt.1[†]						0.99 (0.65)	7.23 (0.63)	8.22 (0.51)	0.12 (0.08)	0.04
Gest	Alt.1[†]						1.29 (0.51)	4.45 (0.43)	5.74 (0.37)	0.22 (0.08)	<0.001
Mum[¥]	Alt.2	-	0.11 (0.01)	0.11 (0.01)	-	-	0.02 (0.03)	0.56 (0.04)	0.58 (0.03)	0.04 (0.05)	0.19
Still[¥]	Alt.2	-	0.18 (0.01)	0.18 (0.01)	-	-	-	0.35 (0.02)	0.35 (0.02)	-	-
Dead[¥]	Alt.2	-	0.25 (0.01)	0.25 (0.01)	-	-	5×10^{-4} (0.03)	0.55 (0.04)	0.55 (0.03)	1×10^{-3} (0.05)	1
Alive	Alt.2	0.02 (0.07)	1.49 (0.1)	1.5 (0.08)	0.01 (0.05)	0.54	0.03 (0.57)	10.91 (0.82)	10.94 (0.67)	3×10^{-3} (0.05)	0.77
Tof	Alt.2	1.55 (0.57)	6.56 (0.53)	8.11 (0.48)	0.19 (0.07)	<0.001	0.67 (0.55)	7.67 (0.62)	8.34 (0.52)	0.08 (0.06)	0.05
Fmor	Alt.2	-	0.01 (8×10^{-4})	0.01 (8×10^{-4})	-	-	-	0.09 (0.01)	0.09 (0.01)	-	-
Wean	Alt.2	-	2.3 (0.12)	2.3 (0.12)	-	-	-	5.14 (0.3)	5.14 (0.3)	-	-
Gest	Alt.2	0.47 (0.17)	1.93 (0.15)	2.4 (0.14)	0.2 (0.06)	<0.001	1.25 (0.49)	4.39 (0.42)	5.64 (0.37)	0.22 (0.08)	<0.001

[¥]Transformed log(trait + 1).[†]Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Table A.6 – Farm 1 Variance Components and Heritability Estimated Using the G Matrix with Single Record Data for Non-Epidemic Phase and Epidemic Phase

		Non- Epidemic Phase					-Epidemic phase				
Trait	Model	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum[¥]	Basic	-	0.11 (0.01)	0.11 (0.01)	-	-	0.04 (0.04)	0.65 (0.05)	0.7 (0.04)	0.06 (0.06)	0.09
Still[¥]	Basic	2×10^{-3} (0.01)	0.18 (0.01)	0.18 (0.01)	0.01 (0.05)	0.6	0.08 (0.04)	0.38 (0.03)	0.46 (0.03)	0.16 (0.07)	<0.001
Dead[¥]	Basic	-	0.25 (0.01)	0.25 (0.01)	-	-	0.05 (0.04)	0.68 (0.05)	0.72 (0.04)	0.07 (0.06)	0.08
Alive	Basic	0.01 (0.07)	1.52 (0.1)	1.53 (0.08)	0.01 (0.05)	0.62	1.26 (0.93)	12.82 (1.04)	14.09 (0.87)	0.09 (0.06)	0.03
Tof	Basic	1.53 (0.57)	6.57 (0.53)	8.11 (0.48)	0.19 (0.07)	<0.001	0.93 (0.6)	7.73 (0.64)	8.66 (0.54)	0.11 (0.07)	0.02
Fmor	Basic	-	0.01 (8×10^{-4})	0.01 (8×10^{-4})	-	-	0.01 (0.01)	0.11 (0.01)	0.12 (0.01)	0.1 (0.06)	0.02
Wean	Basic	-	2.3 (0.12)	2.3 (0.12)	-	-	2.98 (0.88)	6.81 (0.67)	9.79 (0.65)	0.3 (0.08)	<0.001
Gest	Basic	0.49 (0.17)	1.93 (0.15)	2.42 (0.14)	0.2 (0.06)	<0.001	1.5 (0.53)	4.37 (0.43)	5.87 (0.39)	0.26 (0.08)	<0.001
Mum[¥]	Alt.1[†]						0.03 (0.04)	0.65 (0.05)	0.68 (0.04)	0.05 (0.06)	0.15
Still[¥]	Alt.1[†]						0.05 (0.03)	0.39 (0.03)	0.44 (0.03)	0.1 (0.07)	0.03
Dead[¥]	Alt.1[†]						0.01 (0.04)	0.67 (0.05)	0.68 (0.04)	0.02 (0.05)	0.42
Alive	Alt.1[†]						0.57 (0.82)	12.97 (1.01)	13.54 (0.83)	0.04 (0.06)	0.21
Tof	Alt.1[†]						0.72 (0.57)	7.86 (0.64)	8.57 (0.53)	0.08 (0.06)	0.05
Fmor	Alt.1[†]						0.01 (0.01)	0.11 (0.01)	0.12 (0.01)	0.05 (0.06)	0.17
Wean	Alt.1[†]						0.99 (0.65)	7.23 (0.63)	8.22 (0.51)	0.12 (0.08)	0.04
Gest	Alt.1[†]						1.29 (0.51)	4.45 (0.43)	5.74 (0.37)	0.22 (0.08)	<0.001
Mum[¥]	Alt.2	-	0.11 (0.01)	0.11 (0.01)	-	-	0.02 (0.03)	0.56 (0.04)	0.58 (0.03)	0.04 (0.05)	0.19
Still[¥]	Alt.2	-	0.18 (0.01)	0.18 (0.01)	-	-	-	0.35 (0.02)	0.35 (0.02)	-	-
Dead[¥]	Alt.2	-	0.25 (0.01)	0.25 (0.01)	-	-	5×10^{-4} (0.03)	0.55 (0.04)	0.55 (0.03)	1×10^{-3} (0.05)	1
Alive	Alt.2	0.02 (0.07)	1.49 (0.1)	1.5 (0.08)	0.01 (0.05)	0.54	0.03 (0.57)	10.91 (0.82)	10.94 (0.67)	3×10^{-3} (0.05)	0.77
Tof	Alt.2	1.55 (0.57)	6.56 (0.53)	8.11 (0.48)	0.19 (0.07)	<0.001	0.67 (0.55)	7.67 (0.62)	8.34 (0.52)	0.08 (0.06)	0.05
Fmor	Alt.2	-	0.01 (8×10^{-4})	0.01 (8×10^{-4})	-	-	1×10^{-3} (5×10^{-3})	0.09 (0.01)	0.09 (0.01)	0.01 (0.05)	0.49
Wean	Alt.2	-	2.3 (0.12)	2.3 (0.12)	-	-	-	5.14 (0.3)	5.14 (0.3)	-	-
Gest	Alt.2	0.47 (0.17)	1.93 (0.15)	2.4 (0.14)	0.2 (0.06)	<0.001	1.25 (0.49)	4.39 (0.42)	5.64 (0.37)	0.22 (0.08)	<0.001

[¥]Transformed log(trait + 1).[†]Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Table A.7 – Farm 2 Variance Components and Heritability Estimated Using the G Matrix with Single Record Data for Non-Epidemic Phase

Trait	Model	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum[¥]	Basic	0.01 (0.01)	0.1 (0.01)	0.11 (0.01)	0.06 (0.07)	0.2
Still[¥]	Basic	0.01 (0.01)	0.22 (0.02)	0.23 (0.01)	0.04 (0.05)	0.09
Dead[¥]	Basic	0.02 (0.02)	0.24 (0.02)	0.26 (0.02)	0.06 (0.06)	0.1
Alive	Basic	0.04 (0.07)	1.49 (0.11)	1.52 (0.09)	0.02 (0.05)	0.31
Tof	Basic	0.52 (0.56)	8.43 (0.68)	8.95 (0.54)	0.06 (0.06)	0.13
Fmor	Basic	4×10^{-4} (5×10^{-4})	0.01 (6×10^{-4})	0.01 (5×10^{-4})	0.05 (0.06)	0.14
Gest	Basic	0.37 (0.15)	1.21 (0.13)	1.58 (0.1)	0.24 (0.09)	0.01
Mum[¥]	Alt.2	0.01 (0.01)	0.1 (0.01)	0.11 (0.01)	0.06 (0.07)	0.2
Still[¥]	Alt.2	0.01 (0.01)	0.22 (0.02)	0.23 (0.01)	0.05 (0.05)	0.09
Dead[¥]	Alt.2	0.02 (0.02)	0.24 (0.02)	0.26 (0.02)	0.06 (0.06)	0.08
Alive	Alt.2	0.04 (0.08)	1.48 (0.11)	1.53 (0.09)	0.03 (0.05)	0.24
Tof	Alt.2	0.57 (0.58)	8.4 (0.68)	8.97 (0.54)	0.06 (0.06)	0.11
Fmor	Alt.2	4×10^{-4} (5×10^{-4})	0.01 (6×10^{-4})	0.01 (5×10^{-4})	0.05 (0.06)	0.13
Gest	Alt.2	0.35 (0.15)	1.23 (0.13)	1.58 (0.1)	0.22 (0.09)	0.01

[¥]Transformed log(trait + 1).[†]Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Table A.8 – Farm 2 Variance Components and Heritability Estimated Using the G Matrix Repeated Measures Model for Epidemic Including Unknown Phase

Trait	Model	σ^2_A	σ^2_{PE}	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_{PE} LRT P
Mum[¥]	Basic	0.01 (0.02)	2×10^{-7} (2×10^{-8})	0.23 (0.02)	0.25 (0.02)	0.06 (0.08)	0.23	1
Still[¥]	Basic	0.04 (0.03)	1×10^{-7} (1×10^{-8})	0.3 (0.03)	0.34 (0.03)	0.13 (0.09)	0.04	1
Dead[¥]	Basic	5×10^{-3} (0.03)	1×10^{-7} (1×10^{-8})	0.41 (0.04)	0.42 (0.03)	0.01 (0.07)	0.63	1
Alive	Basic	0.23 (0.37)	3×10^{-7} (3×10^{-8})	4.48 (0.45)	4.71 (0.35)	0.05 (0.08)	0.26	1
Tof	Basic	0.87 (0.76)	2×10^{-6} (2×10^{-7})	8.23 (0.85)	9.1 (0.69)	0.1 (0.08)	0.06	1
Fmor	Basic	-	2×10^{-9} (1×10^{-10})	0.02 (2×10^{-3})	0.02 (2×10^{-3})	-	1	-
Gest	Basic	-	0.37 (0.3)	1.81 (0.31)	2.18 (0.16)	-	1	-
Mum[¥]	Alt.1[†]	0.01 (0.02)	2×10^{-7} (2×10^{-8})	0.23 (0.02)	0.25 (0.02)	0.05 (0.08)	0.26	1
Still[¥]	Alt.1[†]	0.04 (0.03)	1×10^{-7} (1×10^{-8})	0.3 (0.03)	0.34 (0.03)	0.12 (0.09)	0.06	1
Dead[¥]	Alt.1[†]	5×10^{-3} (0.03)	1×10^{-7} (1×10^{-8})	0.41 (0.04)	0.42 (0.03)	0.01 (0.07)	0.63	1
Alive	Alt.1[†]	0.24 (0.38)	3×10^{-7} (3×10^{-8})	4.48 (0.45)	4.73 (0.36)	0.05 (0.08)	0.25	1
Tof	Alt.1[†]	0.86 (0.77)	3×10^{-6} (3×10^{-7})	8.27 (0.86)	9.12 (0.69)	0.09 (0.08)	0.07	1
Fmor	Alt.1[†]	-	2×10^{-9} (1×10^{-10})	0.02 (2×10^{-3})	0.02 (2×10^{-3})	-	1	-
Gest	Alt.1[†]	-	0.35 (0.29)	1.79 (0.31)	2.14 (0.16)	-	1	-
Mum[¥]	Alt.2	0.01 (0.02)	2×10^{-7} (2×10^{-8})	0.23 (0.02)	0.24 (0.02)	0.03 (0.08)	0.42	1
Still[¥]	Alt.2	0.04 (0.03)	2×10^{-7} (2×10^{-8})	0.29 (0.03)	0.33 (0.03)	0.13 (0.09)	0.07	1
Dead[¥]	Alt.2	0.01 (0.03)	2×10^{-7} (2×10^{-8})	0.4 (0.04)	0.41 (0.03)	0.01 (0.08)	0.63	1
Alive	Alt.2	0.27 (0.39)	3×10^{-7} (3×10^{-8})	4.33 (0.45)	4.6 (0.35)	0.06 (0.08)	0.23	1
Tof	Alt.2	0.88 (0.79)	2×10^{-6} (3×10^{-7})	8.29 (0.87)	9.16 (0.7)	0.1 (0.08)	0.07	1
Fmor	Alt.2	-	2×10^{-9} (1×10^{-10})	0.02 (2×10^{-3})	0.02 (2×10^{-3})	-	1	-
Gest	Alt.2	-	0.28 (0.29)	1.77 (0.3)	2.05 (0.15)	-	1	-

[¥]Transformed log(trait + 1). - denotes inestimable.

Table A.9 – Joint Farm Variance Components and Heritability Estimated Using the G Matrix with Single Record Data for Non-Epidemic Phase and Epidemic Phase

Trait	Model	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum¥	Basic	5×10^{-3} (4×10^{-3})	0.11 (0.01)	0.11 (5×10^{-3})	0.04 (0.04)	0.12	0.03 (0.03)	0.54 (0.03)	0.57 (0.03)	0.05 (0.05)	0.07
Still¥	Basic	3×10^{-4} (0.01)	0.21 (0.01)	0.21 (0.01)	1×10^{-3} (0.03)	0.8	0.08 (0.03)	0.35 (0.03)	0.43 (0.02)	0.2 (0.07)	<0.001
Dead¥	Basic	3×10^{-4} (0.01)	0.26 (0.01)	0.26 (0.01)	1×10^{-3} (0.03)	1	0.04 (0.03)	0.61 (0.04)	0.65 (0.03)	0.06 (0.05)	0.07
Alive	Basic	1×10^{-3} (0.05)	1.53 (0.07)	1.53 (0.06)	1×10^{-3} (0.03)	0.8	0.65 (0.58)	10.33 (0.68)	10.98 (0.56)	0.06 (0.05)	0.08
Tof	Basic	0.9 (0.4)	7.55 (0.42)	8.45 (0.36)	0.11 (0.05)	<0.001	0.9 (0.52)	7.87 (0.55)	8.77 (0.45)	0.1 (0.06)	0.01
Fmor	Basic	-	0.01 (5×10^{-4})	0.01 (5×10^{-4})	-	-	0.01 (5×10^{-3})	0.09 (0.01)	0.09 (5×10^{-3})	0.09 (0.05)	0.01
Gest	Basic	0.43 (0.11)	1.63 (0.1)	2.06 (0.09)	0.21 (0.05)	<0.001	1.01 (0.34)	3.71 (0.29)	4.71 (0.25)	0.21 (0.07)	9×10^{-5}
Mum¥	Alt.1†						0.03 (0.03)	0.53 (0.03)	0.55 (0.03)	0.05 (0.05)	0.11
Still¥	Alt.1†						0.07 (0.03)	0.35 (0.03)	0.42 (0.02)	0.17 (0.06)	<0.001
Dead¥	Alt.1†						0.03 (0.03)	0.6 (0.04)	0.63 (0.03)	0.05 (0.05)	0.12
Alive	Alt.1†						0.43 (0.54)	10.26 (0.66)	10.7 (0.54)	0.04 (0.05)	0.17
Tof	Alt.1†						0.77 (0.51)	7.96 (0.55)	8.73 (0.45)	0.09 (0.06)	0.03
Fmor	Alt.1†						0.01 (5×10^{-3})	0.08 (0.01)	0.09 (5×10^{-3})	0.07 (0.05)	0.03
Gest	Alt.1†						0.9 (0.33)	3.75 (0.29)	4.65 (0.25)	0.19 (0.07)	<0.001
Mum¥	Alt.2	0.01 (4×10^{-3})	0.1 (0.01)	0.11 (4×10^{-3})	0.05 (0.04)	0.09	0.02 (0.02)	0.46 (0.03)	0.48 (0.02)	0.03 (0.05)	0.19
Still¥	Alt.2	-	0.2 (0.01)	0.2 (0.01)	-	-	0.02 (0.02)	0.34 (0.02)	0.36 (0.02)	0.06 (0.06)	0.12
Dead¥	Alt.2	2×10^{-4} (0.01)	0.25 (0.01)	0.25 (0.01)	9×10^{-4} (0.03)	1	0.01 (0.02)	0.52 (0.03)	0.52 (0.03)	0.01 (0.04)	0.51
Alive	Alt.2	0.01 (0.05)	1.51 (0.07)	1.51 (0.06)	4×10^{-3} (0.03)	0.66	0.13 (0.44)	8.83 (0.56)	8.96 (0.45)	0.01 (0.05)	0.48
Tof	Alt.2	0.95 (0.4)	7.52 (0.42)	8.47 (0.36)	0.11 (0.05)	<0.001	0.76 (0.51)	7.84 (0.54)	8.6 (0.44)	0.09 (0.06)	0.03
Fmor	Alt.2	-	0.01 (5×10^{-4})	0.01 (5×10^{-4})	-	-	2×10^{-3} (3×10^{-3})	0.07 (4×10^{-3})	0.07 (4×10^{-3})	0.03 (0.05)	0.26
Gest	Alt.2	0.43 (0.11)	1.64 (0.1)	2.06 (0.09)	0.21 (0.05)	<0.001	0.87 (0.32)	3.72 (0.28)	4.59 (0.24)	0.19 (0.06)	<0.001

¥Transformed log(trait + 1).†Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Table A.10 – Joint Farm Variance Components and Heritability Estimated Using the G Matrix with Single Record Data for Phase Including Unknown Epidemic

Trait	Model	σ^2_A	σ^2_E	σ^2_P (SE)	h^2 (SE)	σ^2_A LRT P
Mum¥	Basic	0.03 (0.02)	0.53 (0.03)	0.55 (0.03)	0.05 (0.04)	0.09
Still¥	Basic	0.08 (0.03)	0.34 (0.03)	0.42 (0.02)	0.19 (0.06)	<0.001
Dead¥	Basic	0.04 (0.03)	0.6 (0.04)	0.64 (0.03)	0.06 (0.05)	0.08
Alive	Basic	0.67 (0.55)	9.95 (0.64)	10.62 (0.52)	0.06 (0.05)	0.06
Tof	Basic	0.88 (0.5)	7.81 (0.53)	8.68 (0.44)	0.1 (0.06)	0.01
Fmor	Basic	0.01 (5×10^{-3})	0.08 (0.01)	0.09 (4×10^{-3})	0.09 (0.05)	0.01
Gest	Basic	0.87 (0.31)	3.7 (0.27)	4.58 (0.23)	0.19 (0.06)	<0.001
Mum¥	Alt.1†	0.02 (0.02)	0.52 (0.03)	0.54 (0.03)	0.04 (0.04)	0.12
Still¥	Alt.1†	0.07 (0.03)	0.35 (0.02)	0.41 (0.02)	0.16 (0.06)	1×10^{-3}
Dead¥	Alt.1†	0.03 (0.03)	0.59 (0.04)	0.61 (0.03)	0.04 (0.05)	0.13
Alive	Alt.1†	0.44 (0.5)	9.91 (0.62)	10.35 (0.51)	0.04 (0.05)	0.15
Tof	Alt.1†	0.77 (0.5)	7.88 (0.53)	8.66 (0.43)	0.09 (0.06)	0.02
Fmor	Alt.1†	0.01 (4×10^{-3})	0.08 (0.01)	0.09 (4×10^{-3})	0.07 (0.05)	0.03
Gest	Alt.1†	0.78 (0.3)	3.7 (0.27)	4.48 (0.23)	0.18 (0.06)	<0.001
Mum¥	Alt.2	0.01 (0.02)	0.45 (0.03)	0.47 (0.02)	0.03 (0.04)	0.22
Still¥	Alt.2	0.02 (0.02)	0.33 (0.02)	0.35 (0.02)	0.06 (0.05)	0.13
Dead¥	Alt.2	5×10^{-3} (0.02)	0.51 (0.03)	0.52 (0.02)	0.01 (0.04)	0.55
Alive	Alt.2	0.14 (0.42)	8.57 (0.53)	8.71 (0.43)	0.02 (0.05)	0.44
Tof	Alt.2	0.72 (0.49)	7.8 (0.53)	8.52 (0.43)	0.08 (0.06)	0.03
Fmor	Alt.2	2×10^{-3} (3×10^{-3})	0.07 (4×10^{-3})	0.07 (3×10^{-3})	0.03 (0.04)	0.26
Gest	Alt.2	0.78 (0.29)	3.63 (0.26)	4.4 (0.22)	0.18 (0.06)	4×10^{-4}

¥Transformed log(trait + 1).†Alt.1 model not applicable in non-epidemic phase, see basic model results. - denotes inestimable.

Table A.11 – P Values and Effect Size Reported Using GRAMMAR for SNPs Significant at the Chromosome Level for Joint Epidemic Phase Analysis

Trait	Model	SNPS	SSC	Position	P-Value	Effect Size (SE)	MAF
Fmor	Basic	ALGA0121571	3	24,409,351	3×10^{-6}	0.12 (0.03)	0.05
Fmor	Alt.1	ALGA0121571	3	24,409,351	5×10^{-6}	0.09 (0.02)	0.05
Mum	Alt.1	ASGA0018788	4	18,736,505	2×10^{-5}	-0.1 (0.02)	0.14
Mum	Basic	ASGA0031083	7	9,690,196	5×10^{-6}	0.08 (0.02)	0.43
Mum	Alt.1	ASGA0031083	7	9,690,196	7×10^{-6}	0.07 (0.02)	0.43
Mum	Alt.1	ALGA0044856	7	118,565,139	6×10^{-6}	-0.09 (0.02)	0.14
Mum	Basic	MARC0016053	9	135,725,781	2×10^{-7}	0.12 (0.02)	0.14
Fmor	Basic	MARC0016053	9	135,725,781	9×10^{-7}	0.08 (0.02)	0.14
Mum	Alt.1	MARC0016053	9	135,725,781	7×10^{-6}	0.1 (0.02)	0.14
Fmor	Alt.1	MARC0016053	9	135,725,781	4×10^{-7}	0.06 (0.01)	0.14
Still	Basic	ASGA0046584	10	14,519,759	1×10^{-5}	-0.07 (0.01)	0.36
Still	Alt.1	ASGA0046584	10	14,519,759	2×10^{-5}	-0.07 (0.01)	0.36
Still	Basic	ASGA0046593	10	14,647,576	2×10^{-5}	-0.07 (0.01)	0.36
Still	Alt.1	ASGA0046593	10	14,647,576	1×10^{-5}	-0.07 (0.01)	0.36
Still	Alt.2	ASGA0046593	10	14,647,576	2×10^{-5}	-0.09 (0.02)	0.36
Still	Basic	ASGA0053674	12	19,418,067	9×10^{-6}	0.07 (0.02)	0.31
Alive	Basic	ASGA0054360	12	36,158,023	5×10^{-6}	-0.97 (0.21)	0.23
Alive	Alt.1	ASGA0054360	12	36,158,023	2×10^{-6}	-0.95 (0.2)	0.23
Alive	Basic	ALGA0066256	12	38,004,200	5×10^{-6}	-0.92 (0.2)	0.25
Alive	Alt.1	ALGA0066256	12	38,004,200	3×10^{-6}	-0.88 (0.19)	0.25
Mum	Alt.1	ALGA0085129	15	50,703,404	7×10^{-6}	-0.09 (0.02)	0.22
Still	Basic	ALGA0087227	15	138,208,768	3×10^{-6}	-0.09 (0.02)	0.22
Still	Alt.1	ALGA0087227	15	138,208,768	1×10^{-6}	-0.09 (0.02)	0.22
Dead	Alt.1	ALGA0087227	15	138,208,768	9×10^{-6}	-0.11 (0.03)	0.22
Still	Basic	ALGA0087207	15	138,593,445	7×10^{-7}	-0.09 (0.02)	0.23
Still	Alt.1	ALGA0087207	15	138,593,445	4×10^{-7}	-0.09 (0.02)	0.23
Dead	Alt.1	ALGA0087207	15	138,593,445	1×10^{-5}	-0.11 (0.03)	0.23
Still	Basic	ASGA0070725	15	138,715,882	9×10^{-7}	-0.09 (0.02)	0.22
Still	Alt.1	ASGA0070725	15	138,715,882	5×10^{-7}	-0.09 (0.02)	0.22
Dead	Alt.1	ASGA0070725	15	138,715,882	1×10^{-5}	-0.11 (0.03)	0.22
Still	Basic	MARC0016887	17	25,600,780	1×10^{-7}	0.17 (0.03)	0.06
Still	Alt.1	MARC0016887	17	25,600,780	3×10^{-6}	0.15 (0.03)	0.06
Still	Basic	ALGA0097620	18	27,442,071	2×10^{-5}	0.12 (0.03)	0.09

P-value corrected for λ where $\lambda>1$, Effect size is as reported in the GRAMMAR score and therefore is expected to underestimate the true effect

Table A.12 – P Values and Effect Size Reported Using GRAMMAR SNPs Significant at the Chromosome Level for Joint Farm Epidemic Including Unknown Phase Analysis With the Alt.1 Model.

Trait	SNPS	SSC	Position	P-Value	Effect Size (SE)	MAF
Fmor	ASGA0017245	4	3,325,357	1×10^{-5}	-0.04 (0.01)	0.21
Fmor	MARC0016053	9	135,725,781	3×10^{-6}	0.05 (0.01)	0.15
Alive	ALGA0066256	12	38,004,200	1×10^{-5}	-0.76 (0.17)	0.25
Mum	ALGA0085129	15	50,703,404	2×10^{-5}	-0.07 (0.02)	0.23
Still	ALGA0087227	15	138,208,768	7×10^{-6}	-0.07 (0.02)	0.22
Dead	ALGA0087227	15	138,208,768	1×10^{-5}	-0.09 (0.02)	0.22
Still	ALGA0087207	15	138,593,445	1×10^{-6}	-0.08 (0.02)	0.23
Dead	ALGA0087207	15	138,593,445	4×10^{-6}	-0.1 (0.02)	0.23
Still	ASGA0070725	15	138,715,882	4×10^{-6}	-0.08 (0.02)	0.22
Dead	ASGA0070725	15	138,715,882	1×10^{-5}	-0.09 (0.02)	0.22
Still	MARC0016887	17	25,600,780	3×10^{-6}	0.14 (0.03)	0.06

P-value corrected for λ where λ>1, Effect size is as reported in the GRAMMAR score and therefore is expected to underestimate the true effect. Where the same SNP occurs in multiple traits and/or models simplest trait/model shown. Red denotes SNPS also found in Joint Epidemic Phase Analysis, Orange denotes SNPs unique to Epidemic Inc. Unknown

Table A.13 – Baseline Epidemic Phase Heritability Estimates

Trait	Model	σ^2_A (SE)	σ^2_{PE} (SE)	σ^2_E (SE)	h^2 (SE)	LRT
Mum	Basic	0.04 (0.03)	0.09 (0.08)	0.44 (0.08)	0.06 (0.05)	0.04
Still	Basic	0.09 (0.03)	2×10^{-7} (1×10^{-8})	0.34 (0.02)	0.2 (0.06)	8×10^{-5}
Dead	Basic	0.05 (0.03)	0.26 (0.08)	0.34 (0.07)	0.08 (0.05)	0.03
Alive	Basic	0.93 (0.59)	4.32 (1.32)	5.65 (1.19)	0.09 (0.05)	0.02
Tof	Basic	0.92 (0.51)	2.71 (1.13)	5.15 (1.04)	0.1 (0.06)	0.01
Fmor	Basic	0.01 (0.01)	0.04 (0.01)	0.04 (0.01)	0.11 (0.05)	4×10^{-3}
Mum	Alt.1	0.03 (0.03)	0.02 (0.08)	0.5 (0.08)	0.06 (0.05)	0.06
Still	Alt.1	0.07 (0.03)	4×10^{-7} (2×10^{-8})	0.35 (0.02)	0.16 (0.06)	1×10^{-3}
Dead	Alt.1	0.04 (0.03)	0.14 (0.09)	0.44 (0.09)	0.07 (0.05)	0.05
Alive	Alt.1	0.68 (0.55)	2.73 (1.55)	7.06 (1.49)	0.07 (0.05)	0.05
Tof	Alt.1	0.71 (0.49)	2.78 (1.14)	5.18 (1.06)	0.08 (0.06)	0.03
Fmor	Alt.1	0.01 (5×10^{-3})	0.03 (0.01)	0.05 (0.01)	0.09 (0.05)	0.01

Baseline heritability estimates using the full SNP data (including unmapped loci not included in RHM) and full phenotype data (including repeated records not included in RHM) fitting a permanent environmental effect

Abstracts

The following are published conference paper abstracts as detailed on page xxv:

Orrett, C.M., Matika, O., Archibald, A., Lewis, C.R.G., Mclaren, D., Deeb, N. & Bishop, S. (2013). Genetic of host response to infection with porcine reproductive and respiratory syndrome virus (PRRSv). Advances in Animal Bioscience (Proceedings of the British Society of Animal Science 2013, Nottingham). 4. p. 81.

Orrett, C.M., Deeb, N., Pong-Wong, R., Matika, O., Lewis, C.R.G., McLaren, D.G., Archibald, A. & Bishop, S. (2014). Regional Heritability Mapping of Production Traits in Epidemic Porcine Reproductive and Respiratory Syndrome. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. 2014, Vancouver, p. 100.

081

Genetics of host response to infection with porcine reproductive and respiratory syndrome virus (PRRSv)

C M Orrett[1], O Matika[1], A L Archibald[1], C R G Lewis[2], D McLaren[3], N Deeb[2], S C Bishop[1]
[1]The Roslin Institute, Edinburgh, UK, [2]Genus/PIC, Hendersonville, Tennessee, USA, [3]Genus PLC, DeForest, Wisconsin, USA *Email:christopher.orrett@roslin.ed.ac.uk*

Introduction Porcine Reproductive and Respiratory Syndrome (PRRS) is an important viral disease of pigs. Mummification of piglets and other reproductive losses are a direct outcome of PRRS alongside other clinical signs such as respiratory disease, increases in co-infections and occasionally blue colouration on the ears, vulva or hind (Zimmerman *et al.* 2003). Compromised production alongside increased monitoring and treatment costs contribute to a recent cost estimate of €126 per sow per PRRS epidemic in Europe (Nieuwenhuis *et al.* 2012). One option for mitigation of PRRS effects is the exploitation of host genetics, as part of a multifaceted solution incorporating management strategies. Heritable variation in response to the virus has been demonstrated both *in-vivo* and *in-vitro*, reviewed in Lewis *et al.* (2007), indicating the potential for the genetic improvement of host resistance to PRRS. This study aims to investigate the heritability of PRRS resilience in a commercial sow herd using both pedigree and genetic marker information.

Material and methods Data were available from a Chinese commercial multiplication unit over 83 months for 1,821 sows covering 7,907 farrowing events and 87,445 total piglet births. The herd experienced repeated PRRSv outbreaks and data from the farm were partitioned into two groups; epidemic and non-epidemic. Partitioning was done on the basis of trend analysis investigating the rolling 30 day average reproductive traits recorded on farm, as per the threshold-threshold method used by Lewis *et al.* (2009). The available reproductive traits were explored for their ability to identify epidemics; only traits which identified the three, ELISA confirmed outbreaks, were used to partition the data. An initial baseline phase was identified within time periods not exhibiting signs of PRRS for any of the traits; the 95th percentile of this initial baseline was used to identify time windows breaching this threshold. The data within epidemic phase and the remaining non-epidemic phase were used to estimate variance components and heritabilities for the different reproductive outcomes. Pedigree information was available for the sows over 11 generations including 4,249 animals. 60k SNP chip data (Illumina) were available for 637 animals. Traits analysed include mummified piglets per litter, total dead per litter, dead as a fraction of total litter size and gestation length. Restricted Maximum Likelihood (REML) heritability estimates for each trait were generated using ASReml 3.0 for Linux, analysing epidemic and non-epidemic phase farrowing events separately. The significance of fixed effects (Line n=9/Parity n=10) was assessed using the conditional Wald F-Statistic. Random genetic effects were assessed fitting an A matrix (pedigree) in the first instance and a G Matrix (genomic kinship). A Genome Wide Association (GWA) analysis was conducted using the R based GenABEL package to fit the GRAMMAR method.

Results Parity was consistently significant in the models used ($P<0.001$) with differences up to 3.49 piglets between parities for total born dead. Line, whilst not significant, was retained in the model to account for breed differences, with differences up to 3.59 dead piglets. The heritability estimates are shown in Table 1. Total dead piglets showed a 9-fold increase in phenotypic variation from the non-epidemic to epidemic phase, with the increase in genetic variance being even larger, leading to an increased heritability. Similar changes were seen in the other reproductive failure traits, which were all markedly more heritable during epidemic phase. Conversely the heritability of gestation length decreased during the epidemic phase. Using the G matrix altered the heritabilities, however the dataset size was reduced. GWA analysis found chromosome-wide significance on chromosome 4 (SSC4) for the trait of total piglets born dead.

Table 1 Heritability estimates for reproductive traits affected by PRRS

Trait	Non-epidemic phase:	Epidemic phase:	
	A Matrix	A Matrix	G Matrix
Mummified Piglets	0.02 ± 0.01	0.20 ± 0.06	0.09 ± 0.05
Total Dead Piglets	0.05 ± 0.01	0.23 ± 0.06	0.16 ± 0.06
Dead / Total Litter Size	0.03 ± 0.01	0.32 ± 0.06	0.21 ± 0.06
Gestation Length	0.48 ± 0.02	0.23 ± 0.06	0.42 ± 0.05

Conclusions We observed that lowly heritable traits, associated with PRRS outcome, increase in heritability under PRRS challenge. Supported by the increases in genetic variance, this indicates different host control of these outcomes, in the presence of PRRS as opposed to non-epidemic situations. Conversely, gestation length becomes less heritable during PRRS epidemics due to increased phenotypic variance. The significant effect of parity corroborates an earlier finding by Lewis *et al.* (2009), however it is not yet understood whether parity is a proxy for true age effects or risk of previous exposure. The presence of a significant association on SSC4 supports previous findings by Boddicker *et al.* (2012) who also identified a region on SSC4 conferring resistance/susceptibility to PRRSv in growing pigs. These findings indicate that such genetic effects may be translated into potential reproductive benefits, i.e. reductions in mummified and stillborn outcomes. Further work, such as principle component analysis, is required to fully disentangle and quantify the line effects. Current analyses are single breed populations to verify the results from this study.

Acknowledgements The authors gratefully acknowledge funding from Genus plc / PIC and the BBSRC.

References Nieuwenhuis, N., Duinhof, T.F. and Van Nes, A. 2012. Veterinary Record, 170, 225.
Lewis, C. R. G., Ait-Ali, T., Clapperton, M., Archibald, A. L. & Bishop, S. 2007. Viral Immunology, 20, 343-357.
Lewis, C.R.G., Torremorell, M., Galina-Pantoja, L. and Bishop, S.C. 2009. Journal of Animal Science, 87, 876-884.
Boddicker, N., Waide, E.H., Rowland, R.R.R, Lunney, J.K., et al. 2012. Journal of Animal Science, 90, 1733-46.
Zimmerman, JJ, Thacker, BJ and Halbur, PG. 2003. 2003 PRRS Compendium: 2nd Ed, Des Moines, Iowa, National Pork Board.

Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Regional Heritability Mapping of Production Traits in Epidemic Porcine Reproductive and Respiratory Syndrome

C. M. Orrett[1], N. Deeb[2], R. Pong-Wong[1], O. Matika[1], C. R G. Lewis[3], D. G. McLaren[4], A. L. Archibald[1], S. C. Bishop[1]
[1]The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK [2]Genus plc. Hendersonville, Tennessee, USA
[3]PIC Europe, Alpha Building, London Rd, Nantwich UK [4]Genus plc. DeForest, Wisconsin, USA

ABSTRACT: Porcine Reproductive and Respiratory Syndrome (PRRS) is an important disease of pigs. Sow farrowing and service data were obtained from two commercial pig multiplication units which experienced several confirmed PRRS outbreaks. Genomic regions associated with reproductive failure during PRRS outbreaks were investigated using a regional heritability mapping (RHM) approach combining the two datasets. Covariates were explored both fitting and ignoring the shape of the epidemic. Heritability (h^2) of farrowing mortality (FMOR, proportion of dead piglets per litter) was 0.084 ignoring the epidemic shape and 0.059 fitting it. The additive genetic variance was non-estimable for the FMOR trait during non-epidemic phase. Two regions were significantly associated with FMOR at the genome-wide level, on Sus scrofa chromosomes (SSCs) 4 and 7, with several other regions approaching significance. A single SNP on SSC4 was significantly associated ($P<0.001$) with FMOR.

Keywords:
Pig
PRRS
Regional Heritability Mapping

Introduction

Porcine Reproductive and Respiratory Syndrome (PRRS) is an infectious disease of pigs caused by the PRRS virus (PRRSV) with huge financial consequences to the pork industry. Reproductive failure, mummification of piglets and increased pre-weaning losses are direct outcomes of PRRSV infections. These losses appear alongside other clinical signs such as respiratory disease, reduced growth rates, increased incidence of co-infection and occasionally blue coloration on the ears, vulva or hind (Zimmerman et al. 2003). Compromised production plus increased monitoring and treatment costs contribute to losses estimated at $664 million p.a. to the US swine industry alone (Holtkamp, et al. 2012).

Control of PPRS continues to be problematic. The efficacy of current vaccines is limited by the rapid evolution of PRRSV and problems with escape mutants. An option for mitigating the effects of PRRS is the exploitation of host genetic variation in resistance or tolerance to infection; this would form part of a multifaceted solution, incorporating management strategies.

Heritable variation in response to the virus has been demonstrated both in-vivo and in-vitro (Lewis et al. (2007). Recent genome-wide association studies (GWAS) have shown genetic associations of resistance type traits and specific regions of the genome (Boddicker et al. 2012, Orrett et al. 2013, Serão et al. 2014). Boddicker et al. (2012) estimated the heritability (h^2) of variation of both weight gain and viral load at 0.3 in a challenge experiment in growing pigs, indicating the potential for the genetic improvement of host resistance to PRRS.

This paper addresses the genetic control of the impacts of PRRS in reproductive sows. We combine reproductive performance data across two farms that experienced multiple PRRS outbreaks, one of which has been previously described by Lewis et al. (2009). The aim is to identify genomic regions common to the two datasets underlying reproductive failure in a PRRS outbreak.

Materials and Methods

Data. Data were available from two independent commercial pig multiplication units, collected from 2001 to 2007 in one farm and 2009 to 2012 in the other. These were combined to study reproductive traits during several confirmed PRRS outbreaks. The herds experienced repeated PRRS outbreaks over the period, confirmed by a commercial ELISA test (IDEXX, 2003, Maine, with sensitivity of 97.4% and specificity of 99.6%.).

These data described different reproductive outcomes, including numbers of live, stillborn and mummified piglets per litter. Also available were, additional information such as sow line, parity of sow, service date and farrowing date in addition to Illumina PorcineSNP60 BeadChip genotypes. We constructed a new trait, farrowing mortality (FMOR), defined as the number of dead piglets as a proportion of total piglets per litter.

Partitioning the data. Data were partitioned into two groups Epidemic (EPI) and Non-epidemic phase (NON). This was done according to the trend partitioning used by Lewis et al. (2009). The trend used was the rolling 30 day average dead piglets per litter, which showed discrete peaks identifying each of the ELISA confirmed outbreaks. Each farm was partitioned into the two phases separately. Within each farm a baseline period was identified when the herd was not exhibiting signs of PRRS. This period was used to calculate a 95th percentile of dead piglets per litter under non PRRS conditions. Confluent periods where the rolling 30 day average for the trait exceeded the 95th percentile, coinciding with the confirmed outbreaks, were defined as EPI phase data. One period on farm 2 showed a large peak over the threshold which did not coincide with a confirmed outbreak, and data from this period were treated as an unidentified epidemic and discarded. Two weeks either side of each identified